

*Tim Voigt*

# Versuchsplanungsmethoden für die experimentelle Modellbildung bestehender mechatronischer Prozesse

*Dissertation*

Schriftenreihe der Arbeitsgruppe  
Mess- und Regelungstechnik – Mechatronik  
Department Maschinenbau

Herausgeber Oliver Nelles

**Impressum**

Prof. Dr.-Ing. Oliver Nelles  
Arbeitsgruppe Mess- und Regelungstechnik  
Department Maschinenbau  
Universität Siegen  
57068 Siegen  
ISSN 2193-0538  
URN urn:nbn:de:hbz:467-26955  
Zugl.: Siegen, Univ., Diss., 2024

**Versuchsplanungsmethoden für die  
experimentelle Modellbildung bestehender  
mechatronischer Prozesse**

DISSERTATION  
zur Erlangung des Grades eines  
Doktors der Ingenieurwissenschaften

vorgelegt von  
Tim Voigt, M.Sc.  
aus Herford

genehmigt durch die Naturwissenschaftlich-Technische Fakultät  
der Universität Siegen  
Siegen 2024



Betreuer und erster Gutachter  
Prof. Dr.-Ing. Oliver Nelles  
Universität Siegen

Zweite Gutachterin  
Prof. Dr.-Ing. Tamara Reinicke  
Universität Siegen

Tag der mündlichen Prüfung  
25.01.2024



# Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Center for Applied Data Science der Hochschule Bielefeld in Zusammenarbeit mit der Arbeitsgruppe Mess- und Regelungstechnik – Mechatronik der Universität Siegen. Ganz herzlich möchte ich mich bei Herrn Prof. Dr.-Ing. Oliver Nelles von der Universität Siegen und Herrn Prof. Dr.-Ing. Martin Kohlhase von der Hochschule Bielefeld bedanken, die mich während dieser Zeit betreut haben. Das entgegengebrachte Vertrauen, die zahlreichen Diskussionen und die große Unterstützung bei meinen Forschungsvorhaben weiß ich sehr zu schätzen.

Frau Prof. Dr.-Ing. Tamara Reinicke möchte ich für das Interesse an meiner Arbeit und die freundliche Übernahme des Korreferats danken. Darüber hinaus bedanke ich mich bei der Arbeitsgruppe Mess- und Regelungstechnik – Mechatronik für den fachlichen Austausch. Bei meinen Besuchen in Siegen habe ich mich stets über die offene und freundliche Atmosphäre gefreut.

Ich bedanke mich bei allen meinen Kollegen für die gemeinsame Zeit und die vielfältige Unterstützung. Mein Dank gilt insbesondere Marvin Schöne, Nico Migenda, Lukas Vollenkemper und Henning Borchard von der Hochschule Bielefeld. Außerdem gilt mein Dank allen Studierenden, die meine Tätigkeit im Rahmen von Studien-, Bachelor- und Masterarbeiten unterstützt haben.

Der Praxisbezug und die experimentellen Ergebnisse in dieser Arbeit wurden in Zusammenarbeit mit der Firma Miele & Cie. KG erarbeitet. Einen besonderen Beitrag hierzu haben Dr. Armin Peter, Martin Kuhn, Kevin Andre Kupczyk, Matthias Brand und Dominik Osthus geleistet, wofür ich mich herzlich bedanken möchte.

Schließlich möchte ich mich bei meiner Familie für die unendliche Geduld und die große Unterstützung bedanken. Besonders meine Frau Judith und meine Kinder Jule und Jonas haben mich motiviert diese Arbeit erfolgreich fertigzustellen.



---

# Inhaltsverzeichnis

<b>Vorwort</b>	<b>III</b>
<b>Symbole und Abkürzungen</b>	<b>IX</b>
<b>Kurzfassung</b>	<b>XV</b>
<b>Abstract</b>	<b>XVII</b>
<b>1 Einführung</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Stand der Technik . . . . .	4
1.3 Zielsetzung und Aufbau der Arbeit . . . . .	6
<b>2 Grundlagen der experimentellen Modellbildung</b>	<b>11</b>
2.1 Beurteilung von Verfahren zur datenbasierten Modellbildung . . . . .	11
2.1.1 Bewertung der Modellgüte . . . . .	12
2.1.2 Bewertung des Versuchsaufwands . . . . .	15
2.2 Regressionsmodelle . . . . .	16
2.2.1 Polynommodelle . . . . .	17
2.2.2 Lokale Modellnetze . . . . .	19
2.2.3 Gaußprozessmodelle . . . . .	22
2.3 Beurteilung der Eingangsraumabdeckung . . . . .	25
2.3.1 Distanzen zwischen den Versuchspunkten . . . . .	27
2.3.2 Nächste-Nachbarn-Distanz eines raumfüllenden Versuchsplans	28
2.3.3 Kerndichteschätzung und Kullback-Leibler-Divergenz . . . . .	29
2.4 Versuchsplanungsmethoden . . . . .	34
2.4.1 Modellbasierte Versuchspläne . . . . .	34
2.4.2 Modellfreie Versuchspläne . . . . .	36
2.4.3 Sequentielle Versuchspläne . . . . .	39
2.5 Zusammenfassung . . . . .	40

<b>3</b>	<b>Inkrementelle Modellbildung</b>	<b>41</b>
3.1	Einführung und Idee . . . . .	41
3.2	Methodik . . . . .	44
3.2.1	Inkrementelle Versuchsplanung . . . . .	46
3.2.2	Inkrementelle Modelle . . . . .	48
3.3	Modellstrukturen . . . . .	51
3.3.1	Polynommodelle . . . . .	55
3.3.2	Lokale Modellnetze . . . . .	59
3.3.3	Gaußprozessmodelle . . . . .	67
3.4	Beurteilung der Methodik . . . . .	73
3.4.1	Bewertung der Modellgüte und -komplexität . . . . .	73
3.4.2	Berücksichtigung von Prozesswissen . . . . .	77
3.4.3	Praktische Anwendbarkeit . . . . .	79
3.5	Zusammenfassung . . . . .	81
<b>4</b>	<b>Versuchsplanung für zusammengesetzte Einheiten</b>	<b>85</b>
4.1	Einführung . . . . .	86
4.2	Versuchsplanungsmethodik . . . . .	88
4.2.1	Lokaler Suchalgorithmus . . . . .	92
4.2.2	Globaler Suchalgorithmus mittels genetischer Optimierung . . . . .	94
4.2.3	Vergleich der Optimierungsalgorithmen . . . . .	98
4.3	Verbesserte Schätzung von Rauscheinflüssen . . . . .	100
4.4	Reduzierte Versuchspläne mittels Punktselektion . . . . .	103
4.5	Anwendung auf Montageprozesse mit Rotation von Komponenten . . . . .	106
4.5.1	Globaler Versuchsplanungs- und Modellierungsansatz . . . . .	108
4.5.2	Lokaler Versuchsplanungs- und Modellierungsansatz . . . . .	111
4.6	Zusammenfassung . . . . .	114
<b>5</b>	<b>Anwendungen</b>	<b>117</b>
5.1	Inkrementelle Modellbildung einer Heizstrecke . . . . .	117
5.1.1	Anwendung der inkrementellen Modellbildung . . . . .	119
5.1.2	Ergebnisse . . . . .	121
5.2	Modellbasierte Optimierung eines Bitumenofens . . . . .	124
5.2.1	Anwendung der inkrementellen Modellbildung . . . . .	125
5.2.2	Implementierung der Optimierungsstrategie . . . . .	129
5.3	Qualitätsvorhersage für die Montage von Waschmaschinentrommeln . . . . .	134
5.3.1	Lokaler Ansatz zur Vorhersage von Höhenabweichungen . . . . .	136

5.3.2	Globaler Ansatz zur Vorhersage von Rundlaufabweichungen	141
5.4	Zusammenfassung . . . . .	146
<b>6</b>	<b>Zusammenfassung und Ausblick</b>	<b>149</b>
<b>A</b>	<b>Anhang</b>	<b>153</b>
A.1	Grundlagen des EDLS-Algorithmus . . . . .	153
A.2	Berücksichtigung von Randbedingungen im EDLS-Algorithmus . . .	155
A.3	Grundlagen genetischer Optimierung . . . . .	156
	<b>Literaturverzeichnis</b>	<b>161</b>



# Symbole und Abkürzungen

## Allgemeine Symbole

$\epsilon$	Fehlerschwelle
$\underline{\theta}$	Parametervektor
$\mu$	Mittelwert
$\underline{\mu}$	Erwartungswertvektor
$\mu(\cdot)$	Zugehörigkeitsfunktion
$\sigma$	Standardabweichung
$\sigma_f$	Signal-Standardabweichung
$\sigma_n$	Standardabweichung des Rauschens
$\underline{\Sigma}$	Kovarianzmatrix
$\Phi(\cdot)$	Gültigkeitsfunktion
$A$	Anzahl an Individuen
$b$	Modellparameter
$d$	Euklidische Distanz
$d_{\min}$	Minimale Punktdistanz in einem Datensatz
$d_S$	Kriterium zur Beurteilung der größten Datenlücke
$D_{KL}$	Gleichförmigkeitskriterium auf Basis der Kullback-Leibler-Divergenz
$\mathcal{D}$	Menge an Daten- / Versuchspunkten
$e$	Fehler
$\mathcal{E}$	Menge an ausgeschlossenen Versuchspunkten
$f(\cdot)$	Mathematische Funktion
$F(\cdot)$	Fitnessfunktion
$g$	Exponent der Anzahl an Punkten in einer pseudozufälligen Sequenz
$g(\cdot)$	Lokales Modell
$\mathcal{G}$	Menge an Individuen in einer Generation
$i$	Allgemeiner Index

$\underline{I}$	Einheitsmatrix
$j$	Allgemeiner Index
$k$	Zeitindex diskreter Größen / Allgemeiner Index
$\underline{k}$	Kovarianzvektor
$\underline{K}$	Kovarianzmatrix eines Gaußprozessmodells
$K(\cdot, \cdot)$	Kernel-Funktion
$l$	Allgemeiner Index
$L$	Anzahl an lokalen Modellen
$m$	Grad eines Polynoms
$M$	Gesamtzahl an Eingangsgrößen / Dimensionen
$\mathcal{M}$	Menge an möglichen Tauschpartnern
$n$	Rauschen
$N$	Anzahl an Datenpunkten
$p$	Dimensionalität des betrachteten Eingangsraums
$p_K$	Rekombinationswahrscheinlichkeit
$p_M$	Mutationswahrscheinlichkeit
$p_P$	Reproduktionswahrscheinlichkeit
$p(\cdot)$	Wahrscheinlichkeitsdichtefunktion
$P(\cdot)$	Diskrete Wahrscheinlichkeitsverteilung
$\underline{Q}$	Gewichtungsmatrix
$R$	Anzahl an Regressoren
$S$	Individuum
$\mathcal{S}$	Raumfüllender Versuchsplan
$t$	Zeit
$\mathcal{T}$	Testdatensatz
$u$	Eingangsgröße
$\underline{u}$	Eingangsgrößenvektor
$W_G$	Anzahl an Generationen
$x$	Regressor
$\underline{X}$	Datenmatrix
$y$	Ausgangsgröße
$\underline{y}$	Ausgangsgrößenvektor
$z$	Koordinate des Zentrums einer Partition

## Spezifische Symbole für die Inkrementelle Modellbildung

$\delta$	Breite einer Nullstellenfunktion
$\Delta d$	Breite eines Stücks der Transformationsfunktion
$\Psi(\cdot)$	Nullstellenfunktion
$B(\cdot)$	B-Spline
$d$	Partitionsbreite
$i$	Versuchsplanungs- und Modellierungsschritt
$\underline{k}$	Knotenvektor eines B-Splines
$\underline{N}$	Vektor der Anzahl an Datenpunkten
$r$	Transformationsstärke
$\bar{u}$	Koordinate eines Arbeitspunkts
$\underline{u}$	Arbeitspunkt
$z(\cdot)$	Transformationsfunktion für einen inkrementellen Versuchsplan

## Spezifische Symbole für die Versuchsplanung für zusammengesetzte Einheiten

$\alpha$	Rotationswinkel von Komponenten
$c$	Einzelkomponente
$\mathcal{C}$	Menge an Komponenten eines Typs
$I$	Anzahl an verschiedenen Komponententypen
$K$	Anzahl an näherungsweise Wiederholpunkten
$m$	Anzahl an Merkmalen pro Komponente
$\mathcal{P}$	Menge an Komponenten die eine Baugruppe bilden
$q$	Betrachtete Individuen in der Tournament Selection
$s$	Tupeldarstellung einer Permutation
$w$	Anzahl diskreter Winkelpositionen

## Spezifische Symbole für die betrachteten Anwendungen

$\beta$	[°]	Drosselklappenwinkel
$\Delta\hat{\vartheta}_B$	[°C]	Korrektur der Bitumenmattentemperatur
$\Delta h$	[mm]	Planlaufabweichung
$\Delta r$	[mm]	Rundlaufabweichung
$\vartheta_B$	[°C]	Bitumenmattentemperatur
$\vartheta_L$	[°C]	Lufttemperatur
$\vartheta_O$	[°C]	Ofentemperatur
$D_L$	[1/h]	Durchsatz an Laugenbehältern
$E_L$	[kWh]	Energie pro Laugenbehälter
$h_L$	[m]	Laugenbehälterhöhe
$\dot{m}_L$	[kg/h]	Luftmassenstrom
$P_H$	[W / %]	Heizleistung
$P_L$	[%]	Lüfterleistung
$r$	[mm]	radiale Abweichung
$v_B$	[m/s]	Bandgeschwindigkeit

## Mathematische Symbole

$\hat{\cdot}$	Geschätzte oder modellierte Größe
$\tilde{\cdot}$	Transformierte / modifizierte Größe
$\cdot^*$	Modifizierte Größe
$\ \cdot\ $	Norm eines Vektors
$(\cdot)^T$	Transponierte einer Matrix

## Indizes

$\Sigma$	Summe
C	Cluster
d	Dummy
D	Datensatz
G	Gleichverteilung

krit	kritisch
min	Minimal
MC	Monte-Carlo
NN	Nächste-Nachbarn
S	Sobol
Sel	Selektion
soll	Sollwert
test	Testdaten
T	Trommel

## Abkürzungen

ARD	Automatic Relevance Determination
DoE	Versuchsplanung (Design of Experiments)
DP	Dummy-Punkt
EDLS	Extended Deterministic Local Search
GA	Genetischer Algorithmus
GPM	Gaußprozessmodell
ILHAD	Incremental Latin Hypercube Additive Design
KLD	Kullback-Leibler-Divergenz
LHC	Latin Hypercube
LMN	Lokales Modellnetz
LOLIMOT	Local Linear Model Tree
MAE	Mean-Absolute-Error
ML	Maschinelles Lernen
NS	Nicht stationär
OFaaT	One-Factor-at-a-Time
PMX	Partially-Matched-Crossover
RMSE	Root-Mean-Square-Error
SPS	Speicherprogrammierbare Steuerung
SR	Stepwise Regression
TM	Theoretisches Modell



# Kurzfassung

Bestehende mechatronische Prozesse müssen stetig verbessert werden, um aktuelle Anforderungen hinsichtlich Flexibilität, Qualität und Effizienz zu erfüllen. Hierfür werden Modelle benötigt, die das Prozessverhalten abbilden und zur Analyse und Optimierung geeignet sind. Zunehmend werden zu diesem Zweck datenbasierte Modelle verwendet, deren Güte in hohem Maße von den verfügbaren Daten abhängig ist. Daten aus dem regulären Betrieb von Prozessen sind häufig auf wenige Arbeitspunkte konzentriert und somit für eine gute Modellbildung ungeeignet.

Diese Arbeit befasst sich mit der *experimentellen Modellbildung*, bei der die Trainingsdaten eines datenbasierten Modells gezielt in Experimenten erfasst werden. Dazu werden neue *Versuchsplanungsmethoden* entwickelt, die mit einem geringen Aufwand praktisch eingesetzt werden können. Aufwände werden sowohl in der Versuchsplanung als auch -durchführung reduziert, indem verfügbares Prozesswissen einbezogen und Prozesseigenschaften ausgenutzt werden.

Mit der *inkrementellen Modellbildung* wird ein schrittweises Verfahren zur Versuchsplanung und Modellbildung vorgestellt. In jedem Schritt wird eine weitere Eingangsgröße zum Versuchsplan und Modell hinzugefügt, was durch die ansteigende Komplexität zu einem intuitiven Vorgehen führt. Indem die bedeutendsten Eingangsgrößen zuerst betrachtet werden, kann das wesentliche Prozessverhalten in wenigen Schritten abgebildet werden. Darüber hinaus wird eine neue Versuchsplanungsmethode für *zusammengesetzte Einheiten* beschrieben. Bei dieser Art von Prozessen sind die Werte der Eingangsgrößen durch die in den Einheiten verwendeten Komponenten festgelegt. Durch die Ermittlung von optimalen Kombinationen gegebener Komponenten werden kosteneffiziente Versuche z. B. für Montageprozesse ermöglicht.

Die entwickelten Methoden werden anhand von Testprozessen mit Referenzmethoden verglichen. Hierbei werden trotz der gesteigerten praktischen Anwendbarkeit gleichwertige Modellgüten erreicht. Anhand von drei verschiedenen realen Anwendungen wird die praktische Anwendbarkeit demonstriert.



# Abstract

Existing mechatronic processes have to be continuously improved in order to meet current requirements for flexibility, quality and efficiency. For this purpose, models are needed that represent the process behavior and are suitable for analysis and optimization. More and more, data-based models are used for this purpose, whose quality strongly depends on the available data. Data from regular process operation are often concentrated on a few operating points and are therefore not well suited for modeling.

In this work, experimental modeling is applied, in which the training data for a data-based model are systematically obtained in experiments. For this purpose, new *design of experiments methods* are developed that can be practically applied with little effort. Both in the design of experiments and in the execution of experiments, efforts are reduced by incorporating existing process knowledge and by exploiting process characteristics.

With *incremental modelling*, a stepwise approach for experimental design and modelling is presented. In each step, the experimental design and the process model are extended by an additional input variable, which leads to an intuitive modeling approach due to the increasing complexity. By incorporating the most important input variables first, the main process behavior can be modeled in a few steps. Furthermore, a new experimental design method for *composite units* is described. In this type of processes, the values of the input variables are specified by the components used in the units. By creating optimal combinations of given components, cost-efficient experiments are enabled, e.g. for assembly processes.

The developed methods are compared to reference methods using test processes. With these novel methods, equivalent model quality can be achieved even though greater practical applicability is given. The practical applicability is demonstrated by the means of three different real-world applications.



# 1 Einführung

Aktuelle Krisen wie die Rohstoffknappheit in Folge der Corona-Pandemie [1] oder die Energiemangel durch den russischen Angriffskrieg gegen die Ukraine [108] stellen große Herausforderungen für Unternehmen dar. Ein wichtiger Beitrag zur Bewältigung dieser Krisen kann durch einen effizienten Einsatz von Ressourcen und Energie geleistet werden, wozu Produktionsprozesse angepasst werden müssen. Eine Optimierung bestehender Prozesse bietet große Potenziale für Effizienzsteigerungen bei zugleich geringen Investitionskosten und kurzfristigen Möglichkeiten zur Umsetzung [84].

## 1.1 Motivation

Bestehende Prozesse müssen digitalisiert, vernetzt und flexibel ausgestaltet werden, um deren Effizienzpotenziale zu nutzen. Hinzu kommt die Implementierung geeigneter Strategien zur Optimierung der Prozesse hinsichtlich Ressourcen- und Energiebedarf oder weiterer Zielgrößen. Eine Voraussetzung zur praxistauglichen Prozessoptimierung sind *Modelle*, die zuverlässig das benötigte Prozessverhalten abbilden. Mithilfe von Modellen können optimale Betriebsparameter bestimmt werden, ohne den realen Prozess zu manipulieren oder den Produktivbetrieb zu stören. Hierzu kann die in Bild 1.1 dargestellte Struktur zur *modellbasierten Optimierung* eingesetzt werden. Das gewünschte Prozessverhalten wird in Form einer Zielfunktion formuliert. Zusätzlich können Randbedingungen vorgegeben werden, die beispielsweise von der Qualität der eingesetzten Rohstoffe abhängen. In einer Optimierungsschleife werden am Modell optimale Betriebsparameter bestimmt, die anschließend für den Betrieb des realen Prozesses verwendet werden. Durch eine fortlaufende Optimierung der Betriebsparameter können Einflüsse auf den Prozess, beispielsweise durch variierende Rohstoffeigenschaften kompensiert werden und sich ändernde äußere Bedingungen berücksichtigt werden. Eine solche äußere Bedingung kann z. B. die Verfügbarkeit

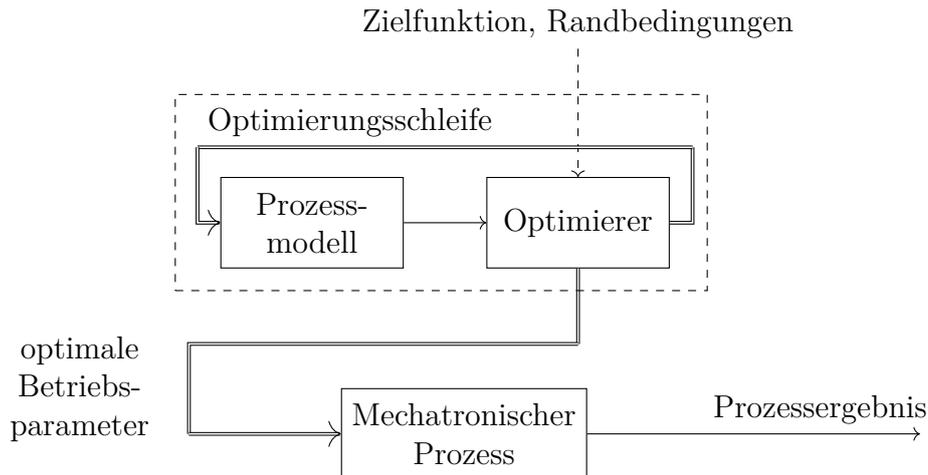


Bild 1.1: Ein gefordertes Prozessergebnis wird durch eine Optimierung am Modell erzielt. Hierzu werden bezogen auf eine Zielfunktion optimale Prozessparameter unter Berücksichtigung von Randbedingungen bestimmt.

von erneuerbaren Energien sein, die in [39] für einen gezielten Betrieb von Prozessen in der Schwerindustrie analysiert wird.

Von vielen bestehenden Prozessen sind keine Modelle aus der Entwicklungsphase vorhanden bzw. existierende Modelle sind durch Änderungen an den Prozessen häufig nicht mehr gültig. Um das wesentliche Prozessverhalten hinsichtlich der definierten Zielgrößen abzubilden, kann sowohl eine theoretische als auch eine datenbasierte Modellbildung erfolgen. Theoretische Modelle werden auf Basis von Gesetzmäßigkeiten wie physikalischen oder chemischen Gleichungen erstellt und können das Prozessverhalten umfassend beschreiben. Allerdings kann bereits eine Modellbildung einzelner Maschinenkomponenten zu einer hohen Komplexität führen, wie es in [45] bei der Modellierung eines Pneumatikzylinders der Fall ist. Bei bestehenden Prozessen erschwert fehlendes Detailwissen über den konstruktiven Aufbau der Prozessbestandteile die theoretische Modellbildung zusätzlich. Datenbasierte Modelle sind hingegen auch bei komplexen Prozessen praktikabel anwendbar und erfordern geringes Prozesswissen. Allerdings ist eine erfolgreiche datenbasierte Modellbildung in hohem Maße von den verwendeten Daten abhängig.

In dieser Arbeit wird sich auf die Modellierung und Optimierung von bestehenden Prozessen bzw. Anlagen fokussiert, wobei eine Form der *experimentellen Modellbildung* gewählt wird. Die Herausforderungen dabei liegen in der Datenqualität und Datenerhebung, für die insbesondere die Verteilung der Datenpunkte im *Eingangsraum* als ausschlaggebend betrachtet wird. Der Eingangsraum wird durch alle im Modell berücksichtigten Eingangsgrößen aufgespannt und beschreibt alle Möglich-

keiten, einen Prozess zu betreiben. Im regulären Prozessbetrieb werden häufig viele Daten gesammelt, die jedoch nur einen gewissen Betriebsbereich repräsentieren. Der Betriebsbereich umschließt typischerweise feste *Arbeitspunkte*, in denen viele Prozesse aktuell betrieben werden. Bezogen auf die Ausdehnung des gesamten Eingangsraums ist der Betriebsbereich häufig klein, was die Gültigkeit eines auf Basis dieser Daten trainierten Modells einschränkt. Die Datengrundlage aus dem regulären Betrieb ist dann nicht ausreichend, um qualitativ hochwertige Modelle abzuleiten, die für eine Prozessoptimierung nutzbar sind.

*Versuchsplanungsmethoden* dienen dazu, mit einem möglichst geringen Versuchsaufwand möglichst viele Informationen über einen Prozess zu erlangen. Hierzu werden Versuchspunkte gezielt im Eingangsraum platziert und anschließend vermessen. In der Praxis werden Versuche häufig nicht systematisch geplant bzw. es werden ungeeignete Methoden wie One-Factor-at-a-Time [21] verwendet. Typische Folgen sind eine geringe Modellgüte und eine ineffiziente Vermessung des Prozesses. Die in dieser Arbeit beschriebenen Methoden zielen auf eine hohe praktische Anwendbarkeit ab. Mehrere Herausforderungen klassischer Versuchsplanungsmethoden wie die Auswahl der betrachteten Eingangsgrößen und die Definition unzulässiger Bereiche im Eingangsraum [29] werden auf diese Weise vereinfacht.

Zusammenfassend ergeben sich die in Bild 1.2 dargestellten Schritte zur Realisierung einer modellbasierten Optimierung für einen bestehenden Prozess. Die Versuchsplanung stellt als einer der ersten Schritte die Grundlage für alle nachfolgenden Schritte dar und hat somit einen entscheidenden Einfluss auf die erfolgreiche praktische Umsetzung. Häufig werden die dargestellten Schritte nicht nur einmalig durchlaufen, sondern bei Bedarf wiederholt. Insbesondere wenn Modelle in der Validierung nicht den Anforderungen genügen, schließen sich neue Versuche an (blauer Pfeil).

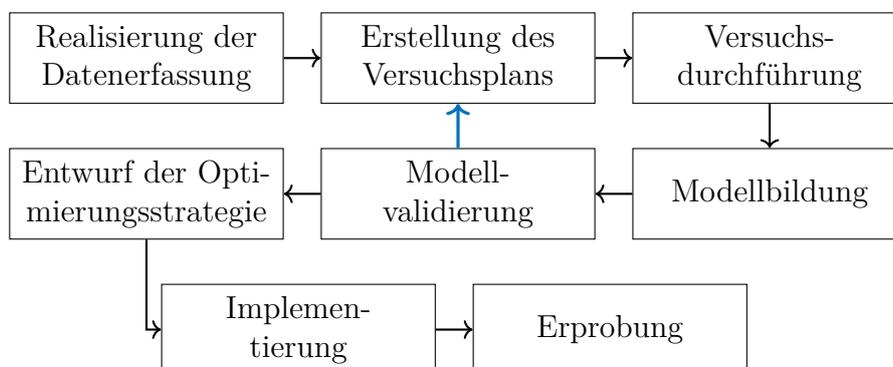


Bild 1.2: Schritte zur Realisierung einer modellbasierten Optimierung. Bei Bedarf werden Schritte wiederholt, z. B. nach der Modellvalidierung (blauer Pfeil).

## 1.2 Stand der Technik

Die experimentelle Modellbildung besitzt ein breites Anwendungsspektrum und ist Gegenstand zahlreicher Arbeiten in Industrie und Forschung. Unter Laborbedingungen sind umfangreiche Möglichkeiten zur Durchführung von Experimenten gegeben. Hierbei stellen klassische Versuchsplanungsmethoden wie *faktorielle Versuchspläne*, *Central-Composite-Designs* oder *Box-Behnken-Designs* in Verbindung mit Polynommodellen einen häufig angewandten Ansatz dar. Beispiele für den Einsatz dieser klassischen Versuchsplanungsmethoden sind bei der Produktentwicklung in der Lebensmittelindustrie [9], bei der Untersuchung von Einflüssen von Biodiesel auf Verbrennungsmotoren [100] oder zur Parameteroptimierung an Werkzeugmaschinen [35, 102] zu finden. Auch komplexe Versuchspläne in Kombination mit flexiblen Modellen sind durch die in Laboren häufig gegebenen Möglichkeiten zur Automatisierung oder Überwachung der Versuche praktikabel einsetzbar. In [53] wird hierzu ein raumfüllender Versuchsplan in Verbindung mit neuronalen Netzen verwendet, um Kräfte und Temperaturen beim Hochgeschwindigkeitsfräsen vorherzusagen. Auch aktives Lernen, bei dem einzelne Versuchspunkte nacheinander auf Grundlage eines fortlaufend aktualisierten Modells platziert werden, wurde erfolgreich zur Entwicklung neuartiger Materialien [65] oder flüssiger Formulierungen [16] angewandt. Schließlich stellt die experimentelle Modellbildung von Verbrennungsmotoren ein typisches Anwendungsgebiet dar, in dem das Prozessverhalten mit dynamischen Modellen abgebildet wird. Hierbei werden geeignete Anregungssignale zur Erzielung einer hohen Modellqualität eingesetzt [11]. Die Erzeugung dieser Anregungssignale kann mithilfe von Versuchsplanungsmethoden erfolgen, wie es in [105] für eine pseudozufällige Sequenz demonstriert wird, die an einen zulässigen Eingangsraum angepasst wird.

Reale Prozesse außerhalb von Laborumgebungen unterliegen üblicherweise einer größeren Anzahl an Einflüssen. Diese Einflüsse können als Eingangsgrößen in die experimentelle Modellbildung einbezogen werden, jedoch ist es häufig nicht oder nur mit einem hohen Aufwand möglich, diese Eingangsgröße auf vorgegebene Werte einzustellen [77]. Zugleich stehen bei produktiv genutzten Prozessen nur eingeschränkte Zeiträume für Versuche zur Verfügung. Auch das Risiko, durch unzulässige Betriebsweisen während der Versuche einen Defekt hervorzurufen, muss bei diesen Prozessen aufgrund hoher Ausfallkosten minimiert werden. Aus diesen Gründen werden für solche Anwendungen zumeist einfach zu planende und zu überwachende Versuchsplanungsmethoden verwendet. Klassische Versuchsplanungsmethoden in Verbindung mit Polynommodellen stellen hierfür ein typisches Vorgehen dar [101], bei dem die

Eingangsgrößen nur auf wenigen Stufen vermessen werden. Praktische Anwendungen sind bei der Parameteroptimierung für einen Abwasserbehandlungsprozess [107] oder bei der Herstellung von Tofu [54] zu finden.

Um die Anwendbarkeit dieser klassischen Versuchsplanungsmethoden zu steigern, sind verschiedene Vorgehensweisen üblich. Sind viele Eingangsgrößen vorhanden, können Screening-Versuchspläne dazu verwendet werden, die wichtigsten Eingangsgrößen zu detektieren. In [7] wird die Anwendung von Screening-Versuchsplänen am Beispiel von Metallschweißprozessen diskutiert. Darüber hinaus kann die Einbeziehung von Prozesswissen beispielsweise in Form physikalischer oder chemischer Gleichungen die experimentelle Modellbildung erleichtern und die Modellgüte steigern. Die resultierenden Modelle werden Gray-Box-Modelle genannt. In [6] wird ein solcher Ansatz zur Erstellung eines Gray-Box-Modells für chemische Prozesse aufgezeigt, der auch in Verbindung mit geplanten Versuchen angewandt werden kann.

Zunehmend werden zur Modellierung real betriebener Prozesse flexible Modellklassen wie neuronale Netze oder Gaußprozessmodelle eingesetzt, die dem Themenfeld des *maschinellen Lernens* (ML) zugeordnet werden können. Allerdings werden diese Modelle häufig auf Basis von Datensätzen erstellt, die durch Prozessbeobachtung im regulären Betrieb erhoben wurden. Hierzu zählen beispielsweise Arbeiten zur Qualitätsvorhersage bei der Lackierung von Autokarosserien mithilfe von neuronalen Netzen [58] oder zur Klassifizierung von Bauteilen beim Metallguss mit einem Random Forest [63]. ML-Modelle von industriellen Prozessen können vom Einsatz von Versuchsplanungsmethoden profitieren. Die Potenziale solcher kombinierten Vorgehensweisen werden in [31] beschrieben, wobei sowohl die Erfassung von Trainingsdaten als auch die Ermittlung von Hyperparametern mittels geplanter Versuche erfolgen kann. In einer realitätsnahen Simulationsstudie werden in [5] die Auswirkungen verschiedener Versuchspläne zur Trainingsdatenerfassung verglichen, wobei mehrere Modellklassen betrachtet werden. Bezogen auf die Modellgüte werden hierbei die besten Ergebnisse mit raumfüllenden Versuchsplänen erzielt.

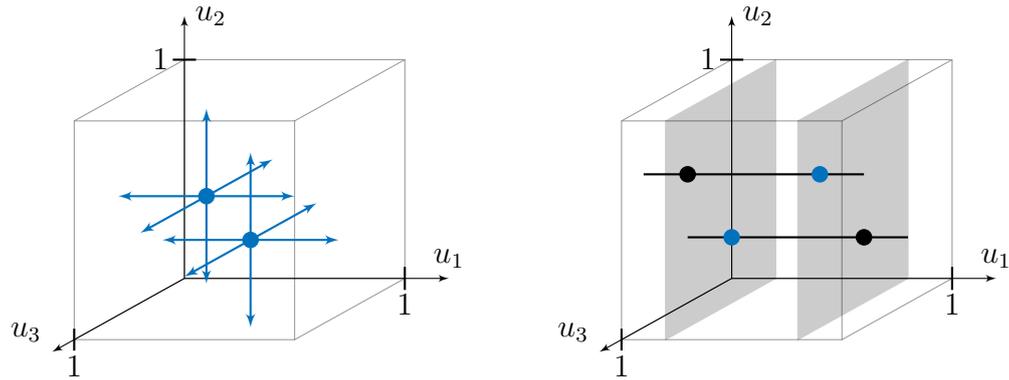
Insgesamt werden Versuchsplanungsmethoden trotz der zahlreichen wissenschaftlichen Arbeiten in diesem Themenfeld häufig nicht in der Praxis angewandt. In [73] wird eine mangelnde Unterstützung der Anwender\*innen bei der Versuchsdurchführung als ein wesentlicher Grund für den Verzicht auf geplante Versuche aufgeführt. Darüber hinaus werden in [86] Anforderungen an Versuchsplanungsmethoden im Kontext von Industrie 4.0 aufgezeigt, zu denen eine geringe Beeinträchtigung der Prozesse und die Erfassung von nichtlinearem Prozessverhalten zählen.

## 1.3 Zielsetzung und Aufbau der Arbeit

Zielsetzung dieser Arbeit ist die Entwicklung von praxistauglichen Methoden, die Versuchsplanung und Modellbildung kombinieren. Diese Methoden sollen den Versuchsaufwand während der experimentellen Modellbildung gering halten und zugleich das Prozessverhalten möglichst im gesamten Eingangsraum erfassen und abbilden. Für das resultierende Modell wird damit ein großer Gültigkeitsbereich angestrebt, der insbesondere für den Einsatz zur modellbasierten Optimierung von Bedeutung ist und eine flexible Bestimmung von Betriebsparametern ermöglicht. Eine erfolgreiche experimentelle Modellbildung ist von dem eingesetzten Prozesswissen abhängig. Entscheidende Aufgaben, die typischerweise auf Grundlage dieses Wissens erfolgen, sind die Definition des Eingangsraums oder die Bewertung der Realisierbarkeit von Datenpunkten. Die Methoden in dieser Arbeit werden auf eine einfache Einbeziehung von Prozesswissen ausgerichtet, um diese Aufgaben gegenüber klassischen Methoden der Versuchsplanung und Modellbildung zu vereinfachen. Zugleich zielen die Methoden auf geringe Aufwände in der Versuchsvorbereitung und -durchführung ab und sollen die Modellbildung und -validierung durch Zusatzinformationen unterstützen. Schließlich ist es ein wesentliches Ziel dieser Arbeit die entwickelten Methoden auf reale Prozesse anzuwenden, um so deren praktischen Nutzen zu beurteilen.

Um diese Ziele zu erreichen, werden zwei verschiedene Verfahren diskutiert, die eine experimentelle Modellbildung ermöglichen, wenn mittels Prozessbeobachtung nur eine unzureichende Datengrundlage erfasst werden kann. Die Auswahl der Verfahren richtet sich nach der Einstellbarkeit der Eingangsgrößen, für die in dieser Arbeit die zwei in Bild 1.3 dargestellten Möglichkeiten betrachtet werden. Einerseits können in (a) die einzelnen Eingangsgrößen unabhängig voneinander auf beliebige Werte eingestellt werden, sodass die Versuchspunkte frei im Eingangsraum platziert werden können. Andererseits sind für das in (b) dargestellte Beispiel nur diskrete Werte der Eingangsgröße  $u_1$  (Ebenen) bzw. feste Kombinationen von  $u_2$  und  $u_3$  (Geraden) zulässig, die durch Eigenschaften von gegebenen Komponenten bestimmt sind. In der Versuchsplanung werden realisierbare Kombinationen an Versuchspunkten (blau) aus allen prinzipiell möglichen Punkten (blau und schwarz) ausgewählt. Zusammenfassend sind der Aufbau dieser Arbeit und die Auswahl des Verfahrens in Bild 1.4 dargestellt.

**Kapitel 2** enthält eine Einführung in die in dieser Arbeit verwendeten Modellklassen und Versuchsplanungsmethoden. Zusätzlich werden Kriterien zur Bewertung der



(a) Freie Einstellbarkeit der Eingangsgrößen für jeden Punkt (b) Einschränkung der möglichen Punkte durch gegebene Komponenten

Bild 1.3: Betrachtete Möglichkeiten der Einstellbarkeit von Eingangsgrößen am Beispiel von zwei Versuchspunkten mit drei Eingangsgrößen  $u_1$ ,  $u_2$  und  $u_3$ .

Eingangsraumabdeckung beschrieben. Diese Kriterien können unter anderem dazu verwendet werden, die Qualität gegebener Datensätze zu beurteilen.

**Kapitel 3** beschreibt mit der inkrementellen Modellbildung ein Vorgehen zur Versuchsplanung und Modellbildung für Prozesse mit frei einstellbaren Eingangsgrößen. Hierbei wird in jedem Schritt eine weitere Eingangsgröße in die Versuchsplanung ein-

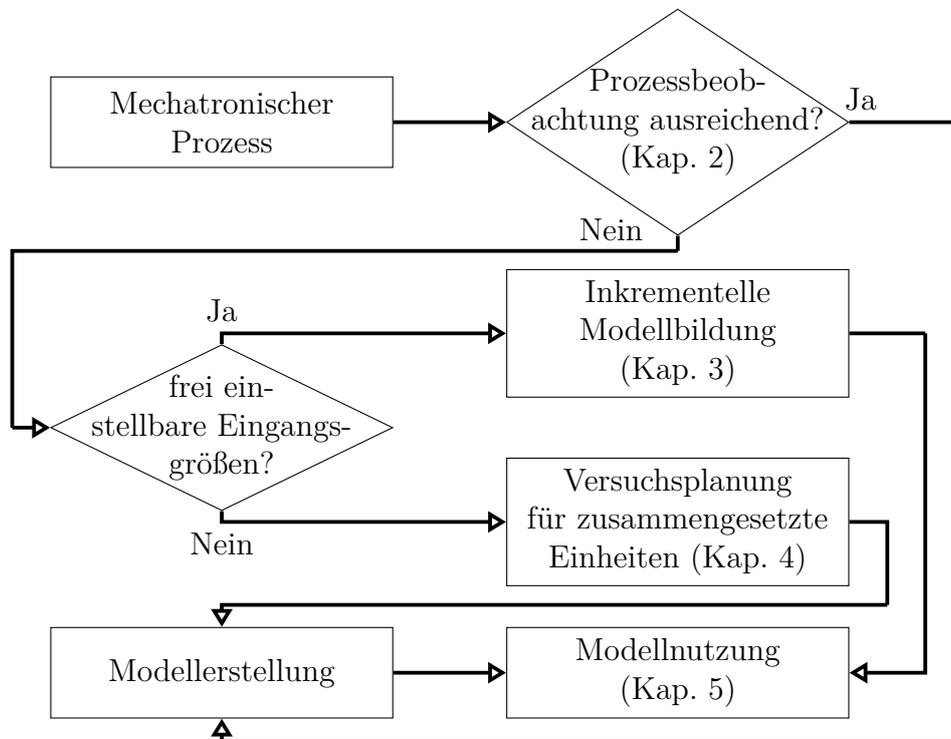


Bild 1.4: Übersicht über den Aufbau dieser Arbeit.

bezogen. Die Einflüsse dieser weiteren Eingangsgröße sowie die Interaktionen mit den zuvor betrachteten Eingangsgrößen werden jeweils durch ein Teilmodell abgebildet. Das Gesamtmodell des betrachteten Prozesses ergibt sich aus der additiven Überlagerung aller Teilmodelle. Für die inkrementelle Modellbildung werden verschiedene Modellklassen betrachtet und die jeweiligen Vor- und Nachteile diskutiert.

**Kapitel 4** betrachtet Prozesse wie beispielsweise Montageprozesse, in denen mehrere Komponenten zu Einheiten zusammengefasst werden. Um diese Prozesse abzubilden, werden spezielle Versuchsplanungsmethoden benötigt, die Einschränkungen durch die verfügbaren Komponenten berücksichtigen. Insbesondere können die einzelnen Eingangsgrößen nicht frei vorgegeben werden, sondern resultieren aus den in einer Einheit verwendeten Komponenten. Mittels verschiedener Optimierungsalgorithmen werden optimale Kombinationen von Komponenten erstellt, die zu einer möglichst gleichmäßigen Abdeckung des Eingangsraums führen. Außerdem werden Maßnahmen zur verbesserten Schätzung von Rauscheinflüssen, zur Reduzierung des Versuchsaufwands mittels Punktselektion und zur Vorgabe der rotatorischen Ausrichtung von Komponenten beschrieben.

**Kapitel 5** präsentiert praktische Anwendungen der entwickelten Verfahren anhand von drei verschiedenen realen Prozessen. An einem Labormodell einer Heizstrecke werden Messungen mit verschiedenen Versuchsplänen und Testdatenverteilungen durchgeführt, um inkrementelle Modelle mit verschiedenen Modellklassen anhand einer realen Anwendung zu vergleichen. An einem Bitumenofen wird die praktische Anwendbarkeit der inkrementellen Modellbildung anhand einer produktiv genutzten Fertigungsanlage demonstriert und eine Optimierungsstrategie auf Basis eines inkrementellen Modells in der Anlagensteuerung implementiert. Schließlich wird die Versuchsplanung für zusammengesetzte Einheiten auf einen Montageprozess von Waschmaschinentrommeln angewandt, wobei zwei verschiedene Ansätze zur Berücksichtigung der Rotation von Komponenten eingesetzt werden.

**Kapitel 6** fasst die vorherigen Kapitel zusammen.

Die wesentlichen Beiträge und Neuerungen dieser Arbeit sind:

- Entwicklung der ILHAD-Methodik zur inkrementellen Modellbildung, die eine additive Modellstruktur und einen entsprechenden Versuchsplan vereint.
- Erweiterung von Polynommodellen, lokalen Modellnetzen und Gaußprozessmodellen für den Einsatz in additiven Modellstrukturen.

- 
- Entwicklung eines Verfahrens zur Anpassung eines inkrementellen Versuchsplans auf Grundlage der LOLIMOT-Partitionierung vorheriger Teilmodelle.
  - Anwendung der inkrementellen Modellbildung zur Verbrauchsreduzierung und Qualitätssteigerung eines Bitumenofens.
  - Weiterentwicklung von Algorithmen (lokaler Suchalgorithmus, genetischer Algorithmus) zur Erstellung von optimalen Versuchsplänen für zusammengesetzte Einheiten.
  - Entwicklung von Strategien zur Punktselektion und zur rotatorischen Ausrichtung von Komponenten bei Montageversuchen.
  - Anwendung der Versuchsplanung für zusammengesetzte Einheiten zur Modellierung von Bauteileinflüssen auf einen Montageprozess für Waschmaschinentrommeln.



## 2 Grundlagen der experimentellen Modellbildung

Die experimentelle Modellbildung zielt darauf ab, ein Prozessverhalten auf Basis von Versuchen abzubilden. Hierzu wurden zahlreiche Methoden entwickelt, die sich für unterschiedliche Arten von Prozessen und Anwendungsgebiete der Modelle eignen [34]. In dieser Arbeit wird das statische Verhalten von Prozessen modelliert, das im Allgemeinen als nichtlinear angenommen wird. Nachfolgend werden die hierfür notwendigen Grundlagen erläutert.

### 2.1 Beurteilung von Verfahren zur datenbasierten Modellbildung

Datenbasierte Modelle sollen ein Prozessverhalten mit einer geforderten Genauigkeit approximieren. Eine höhere Genauigkeit erfordert dabei in der Regel einen größeren Aufwand zur Modellerstellung, wobei in dieser Arbeit die Bereitstellung von Daten als ausschlaggebend für den Aufwand betrachtet wird. In der Praxis muss stets ein Kompromiss zwischen Modellgüte und Modellierungsaufwand gefunden werden, der in Bild 2.1 schematisch dargestellt ist. Werden ausschließlich Daten aus dem regulären Betrieb eines Prozesses verwendet (*Prozessbeobachtung*) ist dieser Aufwand gering und wird hauptsächlich durch die Datenerfassung und -verarbeitung verursacht. Hierfür existieren zahlreiche Standardlösungen, um beispielsweise Daten in eine Cloud-Umgebung zu übertragen [52]. Die so erfassten Daten sind jedoch typischerweise auf wenige Arbeitspunkte konzentriert und besitzen daher einen geringen Informationsgehalt. Dies schränkt den Gültigkeitsbereich eines Modells ein und führt bezogen auf den gesamten Eingangsraum zu einer geringen Modellgüte.

Werden Versuche durchgeführt, um einen Trainingsdatensatz für ein Modell zu erstellen (*experimentelle Modellbildung*) ist der Versuchsaufwand die dominierende Größe.

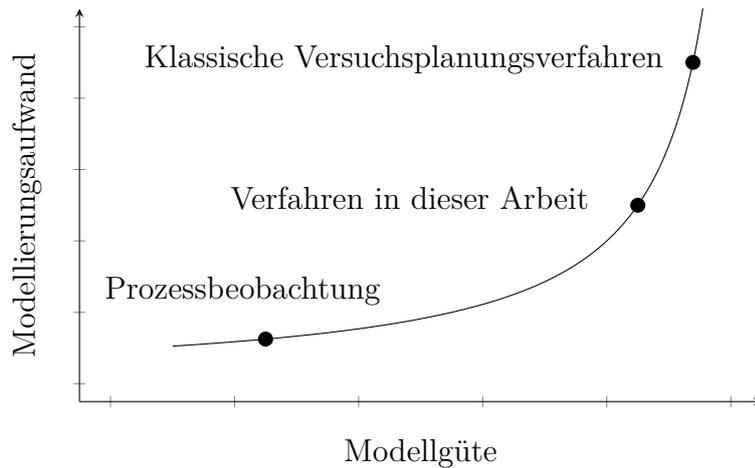


Bild 2.1: Qualitativer Vergleich von Modellgüte und Modellierungsaufwand verschiedener Verfahren zur datenbasierten Modellbildung (in Anlehnung an [114]).

Viele klassische Versuchsplanungsmethoden wie beispielsweise faktorielle Versuchspläne ermöglichen zwar mit wenig Datenpunkten eine hohe Modellgüte jedoch ist die Vermessung der einzelnen Datenpunkte in der Praxis häufig aufwändig. Die in dieser Arbeit beschriebenen Methoden zielen darauf ab, bei nur geringen Einbußen in der Modellgüte den Modellierungsaufwand gegenüber dem Einsatz von klassischen Versuchsplanungsmethoden deutlich zu reduzieren. Möglichkeiten zur Bewertung der Modellgüte und des Versuchsaufwands werden in den nachfolgenden Abschnitten erläutert.

Darüber hinaus existieren weitere Anforderungen, die beispielsweise die Wahl der Modellklasse beeinflussen. Für einen sicheren Betrieb von Modellen wird häufig gefordert, dass das erlernte Prozessverhalten beurteilt werden kann [99]. Hierfür werden interpretierbare Modelle benötigt, bei denen durch einfache Darstellungen oder Analysen Informationen über die vom Modell erlernten Zusammenhänge gewonnen werden können [78]. Weitere Anforderungen wie eine Adaptierbarkeit oder ein maximaler Speicherbedarf können die Auswahl der Modellklasse zusätzlich beeinflussen.

### 2.1.1 Bewertung der Modellgüte

Ein übliches Fehlermaß zur Messung der Modellgüte ist der *Mean-Absolute-Error* (MAE), der sich auf Basis eines Testdatensatzes mit  $N_{\text{test}}$  Datenpunkten zu

$$\text{MAE} = \frac{1}{N_{\text{test}}} \sum_{k=1}^{N_{\text{test}}} |y(k) - \hat{y}(k)| \quad (2.1)$$

ergibt. Hierbei bezeichnet  $y(k)$  mit  $k = 1, 2, \dots, N_{\text{test}}$  die Ausgangsgröße des zu modellierenden Prozesses und  $\hat{y}(k)$  den Modellausgang. Die Abweichung dieser beiden Größen wird als Modellfehler  $e(k) = y(k) - \hat{y}(k)$  bezeichnet. Ein Fehlermaß, das große Abweichungen stärker als der MAE berücksichtigt, ist der *Root-Mean-Squared-Error* (RMSE). Dieser berechnet sich zu

$$\text{RMSE} = \sqrt{\frac{1}{N_{\text{test}}} \sum_{k=1}^{N_{\text{test}}} (y(k) - \hat{y}(k))^2} \quad (2.2)$$

und basiert auf den Quadraten der Modellfehler an den einzelnen Testdatenpunkten. Beide Fehlermaße sind üblich, um die Güte von datenbasierten Modellen zu bewerten und bieten unterschiedliche Empfindlichkeiten bezüglich Ausreißern [17].

Zum Vergleich dieser Fehlermaße wird in Bild 2.2 ein Prozess mit einer Eingangsgröße  $u$  und einem exponentiellen Prozessverhalten durch zwei quadratische Polynommodelle approximiert. Zur Bewertung der beiden Modelle wird ein Testdatensatz mit  $N_{\text{test}} = 8$  Datenpunkten (Sterne) verwendet, der den eindimensionalen Eingangsraum gleichförmig abdeckt. Für beide Modelle ergibt sich ein nahezu identischer MAE, zu dem in (a) alle Testdatenpunkte ähnliche Beiträge leisten, während in (b) der MAE hauptsächlich aus dem Modellfehler des blau markierten Datenpunkts resultiert. Dieser blaue Datenpunkt liegt in einem Bereich des Eingangsraums (orange Markierung), in dem das Modell in (b) große Abweichungen aufweist. Der RMSE

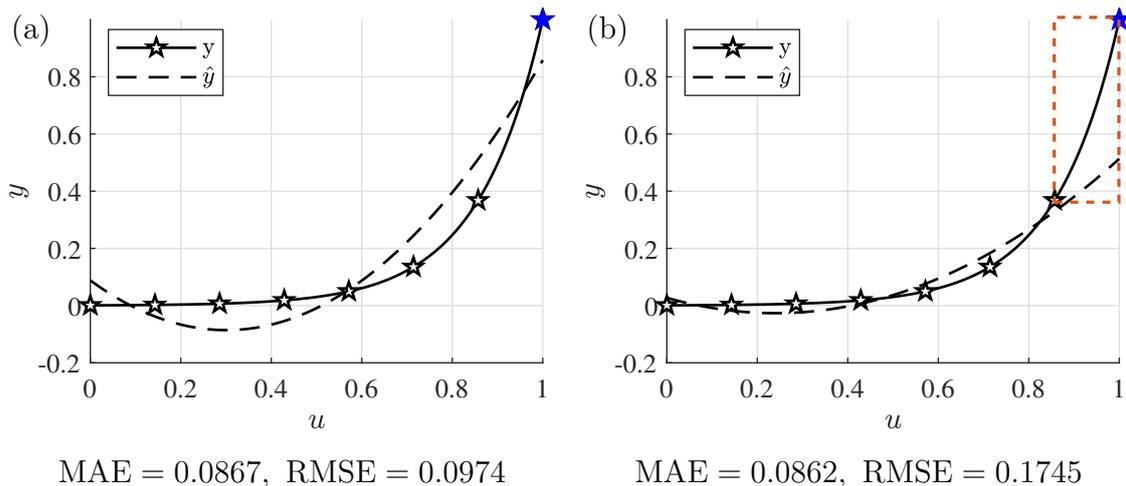


Bild 2.2: Vergleich zweier quadratischer Modelle, die ein exponentielles Prozessverhalten approximieren. Beide Modelle weisen einen vergleichbaren MAE aber unterschiedliche Werte des RMSE auf. Ein für viele Anwendungen kritischer Bereich mit großem Modellfehler ist in (b) hervorgehoben.

wird durch den Modellfehler bei diesem Datenpunkt stark beeinflusst. Da in vielen Anwendungen große Fehler kritisch sind, ist die Sensitivität des RMSE bezüglich großer Modellfehler eine erwünschte Eigenschaft. Dieses Fehlermaß wird daher im Folgenden zur Bewertung und zum Vergleich verschiedener Modelle verwendet.

## Bias-Varianz-Dilemma

In vielen praktischen Anwendungen sind Näherungen wie die in Bild 2.2 dargestellte Approximation einer Potenzfunktion durch ein quadratisches Polynommodell ausreichend. Der Fehler, der hierbei aus der mangelnden Flexibilität des gewählten Polynommodells resultiert, wird als *Biasfehler* bezeichnet. Ein zweiter Anteil des Modellfehlers ist der *Varianzfehler*, der sich aus verrauschten oder ungünstig verteilten Trainingsdaten ergibt und zu Abweichungen der geschätzten Modellparameter führt. Da der Varianzfehler mit steigender Anzahl an Parametern und damit größerer Modellflexibilität zunimmt, während zugleich der Biasfehler abnimmt, muss in der Praxis ein Kompromiss gefunden werden [80]. Die Notwendigkeit zur Abwägung zwischen Bias- und Varianzfehler wird als *Bias-Varianz-Dilemma* bezeichnet.

## Testdatenverteilung

Wird ein Modell wie in Bild 2.2b für  $u \rightarrow 1$  in Bereichen des Eingangsraums mit großen Modellfehlern betrieben, kann dies in der Anwendung zu unerwünschtem Verhalten führen. Um solche Bereiche zu detektieren, ist die Verteilung der Testdaten von Bedeutung. Typischerweise wird hierfür dieselbe Verteilung gefordert wie für die Trainingsdaten. Das übliche Vorgehen einen Datensatz in Trainings- und Testdaten aufzuteilen [80] ist aufgrund der geringen Datenmenge bei Versuchsdaten häufig nicht praktikabel. Stattdessen können Testdatenverteilungen gezielt vorgegeben werden, die für den Betrieb des Modells relevante Bereiche widerspiegeln und mit einem akzeptablen Aufwand erfasst werden können. Je nach Verwendungszweck des Modells bieten sich Datenverteilungen an, die den gesamten Eingangsraum gleichmäßig abdecken oder auf bestimmte Bereiche wie einen typischen Arbeitspunkt konzentriert sind. In dieser Arbeit werden Trainingsdaten gezielt mittels Versuchsplanungsverfahren generiert, wodurch sich im Allgemeinen unterschiedliche Verteilungen der Test- und Trainingsdaten ergeben. Eine umfassende Beschreibung der aus den verschiedenen Datenverteilungen resultierenden Probleme ist in [13] zu finden.

### 2.1.2 Bewertung des Versuchsaufwands

In der Literatur wird häufig die Anzahl  $N$  an zu vermessenen Datenpunkten als ausschlaggebend für den gesamten Versuchsaufwand angenommen [103]. Je nach Eigenschaften des betrachteten Prozesses und dem Vorgehen zur Versuchsdurchführung kann allerdings der Aufwand pro Versuchspunkt in der Praxis stark variieren. Der Versuchsaufwand resultiert aus der Vermessung aller durch einen Versuchsplan vorgegebenen Datenpunkte. Ein solcher Messvorgang kann in mehrere Phasen aufgeteilt werden, die in Bild 2.3 für einen Prozess mit einer Eingangsgröße über die Zeit  $t$  schematisch dargestellt sind. In der Phase i) wird die Eingangsgröße auf den durch den Versuchsplan festgelegten Wert eingestellt. Diese Phase ist besonders lang, wenn Eingangsgrößen durch aufwändige manuelle Tätigkeit wie beispielsweise gezielte Materialzuführung oder Werkzeugwechsel eingestellt werden. Auch die Einstellung von dynamischen Prozessgrößen wie Temperaturen oder Füllständen kann zu einer langen Einschwingzeit führen, bis stationäre Endwerte dieser Größen erreicht sind. Der aus der Einstellung aller Eingangsgrößen resultierende Aufwand wird im Folgenden als Stellaufwand bezeichnet. Dieser Stellaufwand ist beim Vorliegen manueller Tätigkeiten umso größer, je mehr Größen manuell eingestellt werden. In der Phase ii) wird der Prozess mit den eingestellten Eingangsgrößen betrieben, wobei die Dauer dieser Phase für den jeweiligen Prozess charakteristisch ist, aber auch von den Eingangsgrößen abhängen kann. Ausgangsgrößen müssen häufig unmittelbar nach Beendigung des Prozesses erfasst werden, wie es beispielsweise bei einer Bauteiltemperatur notwendig ist. Teilweise kann die Vermessung der Ausgangsgrößen zu einem späteren

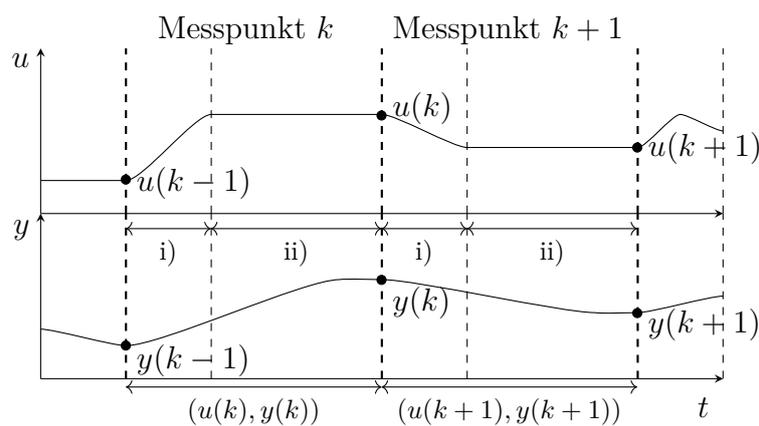


Bild 2.3: Vermessung des statischen Verhaltens eines Prozesses mit einer Eingangsgröße  $u(k)$  und einer Ausgangsgröße  $y(k)$  zu den diskreten Zeitpunkten  $k = 1, 2, \dots, N$ . Für jeden Datenpunkt  $(u(k), y(k))$  wird in i) die Eingangsgröße eingestellt und nach Abschluss von ii) die Ausgangsgröße gemessen.

Zeitpunkt erfolgen, wenn z. B. Maße verschiedener Bauteile als Ausgangsgrößen verwendet werden. Die Anzahl an Datenpunkten in einem Versuchsplan gibt vor, wie viele Zyklen aus Phase i) und Phase ii) in den Versuchen durchlaufen werden.

Neben der Minimierung der Anzahl  $N$  an Versuchspunkten ermöglichen somit praktikable Vorgehensweisen bei der Versuchsplanung und -durchführung eine Reduzierung des Versuchsaufwands. Aus betriebswirtschaftlicher Sicht kann der gesamte Versuchsaufwand in Form von Kosten beschrieben werden, wobei die für Versuche wichtigsten Kostenverursacher nachfolgend aufgeführt sind:

- **Maschinen:** Der betrachtete Prozess wird während der Versuche in verschiedene Betriebspunkte überführt, die zu unzureichenden Prozessergebnissen führen können. Aus diesem Grund steht der Prozess für die Dauer der Versuche nicht für den Produktivbetrieb zur Verfügung. Zusätzlich führen die Versuche zu einer Abnutzung der Produktionsanlage. Die resultierenden Maschinenkosten können in Form eines Maschinenstundensatzes beziffert werden [27].
- **Material:** Für die Durchführung von Versuchen wird ausreichend Material benötigt. Häufig wird hierbei von konstanten Materialeigenschaften ausgegangen bzw. der Einfluss von unterschiedlichen Materialchargen durch Blockbildung kompensiert [77]. Werden Materialeigenschaften mit in die Versuchsplanung einbezogen, steigert dies den Versuchsaufwand. So muss das benötigte Material mit den geforderten Eigenschaften beschafft bzw. hergestellt werden. Dieses Material verursacht somit Beschaffungs- bzw. Herstellkosten [27].
- **Manuelle Tätigkeiten:** Die Durchführung von Versuchen erfordert in der Regel manuelle Eingriffe in einen Prozess. Hierbei müssen beispielsweise Eingangsgrößen manuell eingestellt, Material und Produkte gekennzeichnet oder Messungen durchgeführt werden. Der betrachtete Prozess kann selbst manuelle Tätigkeiten wie z. B. manuelle Montagevorgänge beinhalten. Die hierdurch verursachten Personalkosten tragen zum Versuchsaufwand bei [27].

## 2.2 Regressionsmodelle

In dieser Arbeit wird das statische Prozessverhalten von industriellen Prozessen mithilfe von Regressionsmodellen abgebildet. Die verwendete Struktur zur Modellerstellung ist in Bild 2.4 dargestellt, wobei sowohl der Prozess als auch das Modell nichtlineares Verhalten aufweisen können. Der zu modellierende Prozess besitzt  $M$

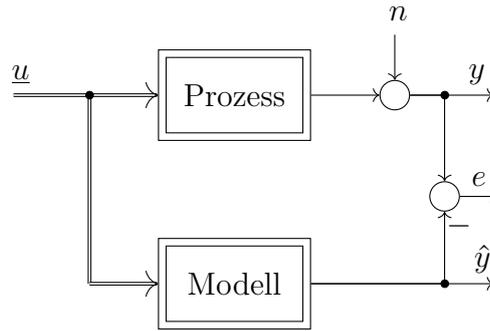


Bild 2.4: Struktur zur Erstellung eines Regressionsmodells [80].

Eingangsgrößen, die zu einem Eingangsgrößenvektor  $\underline{u} = [u_1 \ u_2 \ \dots \ u_M]^T$  zusammengefasst werden. In Abhängigkeit dieses Eingangsgrößenvektors ergibt sich eine Ausgangsgröße  $y$ , die zusätzlich einen stochastischen Anteil durch ein additives Prozessrauschen  $n$  aufweist. Das Prozessmodell trifft mit seinem Modellausgang  $\hat{y}$  Vorhersagen auf Grundlage des Eingangsgrößenvektors  $\underline{u}$ . Im Modelltraining werden die Fehler  $e(k) = y(k) - \hat{y}(k)$  auf Basis der Trainingsdatenpunkte  $\{\underline{u}(k), y(k)\}$  mit  $k = 1, 2, \dots, N$  berechnet. Diese Fehler werden genutzt, um das Modellverhalten an den Prozess anzupassen. Hierzu kann neben den Modellparametern auch die Modellstruktur auf Grundlage des Fehlers angepasst werden.

Es existiert eine Vielzahl an Modellklassen und Algorithmen, die zur Lösung von Regressionsproblemen verwendet werden können. Modelle können unterschieden werden in *parametrische* Modelle mit einer festgelegten Anzahl an Modellparametern und *nichtparametrische* Modelle, bei denen sich die Anzahl der Parameter während des Modelltrainings ergibt [80]. Für diese Arbeit wurden drei häufig verwendete Modellklassen aus beiden Kategorien ausgewählt, die einen unterschiedlichen Grad an Flexibilität bieten. Nachfolgend werden diese Modellklassen genauer beschrieben.

### 2.2.1 Polynommodelle

Eine zur Modellierung von industriellen Prozessen häufig verwendete Modellklasse sind Polynommodelle. Diese besitzen eine einfache Modellstruktur, die zu einer hohen Interpretierbarkeit und einer einfachen Implementierung in technischen Systemen führt. Ein Polynommodell kann als gewichtete Summe

$$\hat{y} = b_0 + \sum_{j=1}^R b_j \cdot x_j \quad (2.3)$$

$R$  verschiedener Regressoren  $x_j$  dargestellt werden, wobei die Modellparameter  $b_j$  mit  $j = 1, 2, \dots, R$  als Gewichte und  $b_0$  als Offset bezeichnet werden. Als Regressoren werden für jede Eingangsgröße  $u_i$  mit  $i = 1, 2, \dots, M$  jeweils alle Potenzen  $u_i, u_i^2, \dots, u_i^m$  bis zu einem vorgegebenen Grad  $m \in \mathbb{N}$  des Polynoms verwendet. Werden Potenzen verschiedener Eingangsgrößen miteinander multipliziert, ergeben sich Regressoren, die Interaktionen zwischen den Eingangsgrößen darstellen [28]. Hierbei werden nur solche Regressoren verwendet, in denen die Summe aller Potenzen den Grad  $m$  des Polynoms nicht übersteigen. Allgemein kann ein Polynommodell des Grades  $m$  mit  $M$  Eingangsgrößen beschrieben werden als:

$$\hat{y} = b_0 + b_1 \underbrace{u_1}_{x_1} + b_2 \underbrace{u_2}_{x_2} + \dots + b_M \underbrace{u_M}_{x_M} + b_{M+1} \underbrace{u_1^2}_{x_{M+1}} + b_{M+2} \underbrace{u_1 u_2}_{x_{M+2}} + \dots + b_R \underbrace{u_M^m}_{x_R}. \quad (2.4)$$

Die Modellparameter des Polynommodells können mit einem geringen Rechenaufwand mittels der *Methode der kleinsten Fehlerquadrate* (engl. *least squares*) bestimmt werden. Hierzu wird die Summe der Fehlerquadrate minimiert

$$\min_{\hat{\theta}} \sum_{k=1}^N (y(k) - \hat{y}(k))^2, \quad (2.5)$$

wobei  $\hat{\theta} = [\hat{b}_0 \ \hat{b}_1 \ \dots \ \hat{b}_R]^T$  der geschätzte Parametervektor ist. Zur analytischen Lösung dieses Minimierungsproblems werden die gegebenen Trainingsdaten wie folgt zu einer Regressionsmatrix  $\underline{X}$  und einem Ausgangsgrößenvektor  $\underline{y}$  zusammengefasst:

$$\underline{X} = \begin{bmatrix} 1 & x_1(1) & x_2(1) & \dots & x_R(1) \\ 1 & x_1(2) & x_2(2) & \dots & x_R(2) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_1(N) & x_2(N) & \dots & x_R(N) \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix}. \quad (2.6)$$

Mittels Optimierung von Gleichung (2.5) kann der geschätzte Parametervektor zu

$$\hat{\theta} = (\underline{X}^T \underline{X})^{-1} \underline{X}^T \underline{y} \quad (2.7)$$

bestimmt werden [80].

In der Anwendung von Polynommodellen stellt die Wahl des Grades  $m$  eine kritische Aufgabe dar. Wird dieser zu gering gewählt, kann das Prozessverhalten nur unzureichend abgebildet werden, sodass ein großer Biasfehler entsteht. Sehr flexible Polynommodelle mit einem hohen Grad können hingegen zu einem hohen Varianzfehler

führen, der sich in unerwünschten Oszillationen und einem schlechten Extrapolationsverhalten äußert [80].

## 2.2.2 Lokale Modellnetze

Als gut interpretierbare Modellklasse, die auch zur Abbildung komplexer funktionaler Zusammenhänge geeignet ist, werden in diesem Abschnitt *lokale Modellnetze* (LMN) beschrieben. LMN unterteilen den betrachteten Eingangsraum in mehrere Partitionen. Jeder dieser Partitionen ist ein lokales Modell zugeordnet, das das Prozessverhalten in dem jeweiligen Bereich des Eingangsraums repräsentiert. Als lokale Modellstrukturen kommen zumeist einfache Modelle, wie affine Modelle (Polynommodelle ersten Grades) zum Einsatz. LMN wurden erfolgreich zur Modellierung des statischen und dynamischen Verhaltens von industriellen Prozessen eingesetzt [49, 89].

Zur Erstellung von lokalen Modellnetzen existieren zahlreiche Algorithmen. Aufgrund der einfacheren Struktur und Interpretierbarkeit wird in dieser Arbeit der LOLIMOT-Algorithmus (Local Linear Model Tree) [79] verwendet. Darüber hinaus existieren Erweiterungen von LOLIMOT, wie beispielsweise der LOPOMOT-Algorithmus (Local Polynomial Model Tree) [96], der eine Selektion von Regressoren bei lokalen Polynommodellen beinhaltet, oder der HILOMOT-Algorithmus (Hierarchical Local Model Tree) [48], bei dem achsenschräge Teilungen des Eingangsraums vorgenommen werden.

Der LOLIMOT-Algorithmus dient zur Unterteilung des Eingangsraums in Partitionen, in denen jeweils ein lokales Modell gültig ist. Eine beispielhafte Partitionierung eines zweidimensionalen Eingangsraums durch den LOLIMOT-Algorithmus ist in Bild 2.5 zu sehen. Die Partitionierung erfolgt ausgehend von einem globalen Modell im ersten Iterationsschritt, dessen Partition den vollständigen Eingangsraum umfasst. In jeder Iteration wird fortlaufend das Modell mit dem größten Modellfehler (grau) mittig und achsenorthogonal geteilt, wobei alle Teilungsmöglichkeiten in zwei gleich große Partitionen überprüft werden. Die Teilung, die zu dem geringsten Fehler des Gesamtmodells führt, wird durchgeführt. LOLIMOT unterteilt den Eingangsraum so lange, bis ein Abbruchkriterium erreicht ist. Üblich sind z. B. eine maximale Anzahl an lokalen Modellen oder ein Schwellenwert für den Trainingsfehler.

Jeder der  $L$  Partitionen in einem Iterationsschritt wird eine Gültigkeitsfunktion  $\Phi_l(\underline{u})$  mit  $l = 1, 2, \dots, L$  und ein lokales Modell  $g_l(\underline{u})$  zugeordnet. Die Gültigkeitsfunktionen gewichten die einzelnen Teilmodelle in Abhängigkeit der betrachteten Position

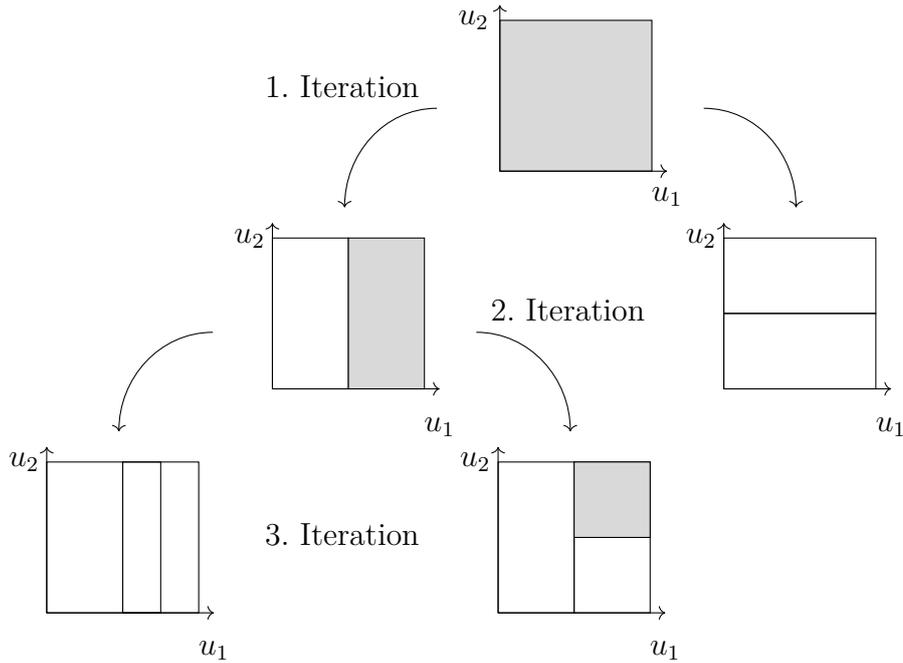


Bild 2.5: Exemplarische Partitionierung eines zweidimensionalen Eingangsraums durch den LOLIMOT-Algorithmus. Inkrementell wird das lokale Modell mit dem größten Modellfehler (graue Markierung) achsenorthogonal und mittig in kleinere Partitionen unterteilt. Die bessere der möglichen Partitionierungen wird realisiert (in Anlehnung an [80]).

im Eingangsraum. Durch die gewichtete Überlagerung aller Teilmodelle ergibt sich der Modellausgang zu:

$$\hat{y} = \sum_{l=1}^L g_l(\underline{u}) \cdot \Phi_l(\underline{u}). \quad (2.8)$$

Die Gültigkeitsfunktionen sollen glatte Übergänge zwischen den einzelnen Teilmodellen ermöglichen, die in vielen Anwendungen vorteilhaft sind. Hierzu werden standardmäßig normierte Gaußfunktionen verwendet [80]. Eine Gaußfunktion

$$\mu_l(\underline{u}) = \exp \left( -\frac{1}{2} \left( \frac{(u_1 - z_{l,1})^2}{\sigma_{l,1}^2} + \dots + \frac{(u_M - z_{l,M})^2}{\sigma_{l,M}^2} \right) \right) \quad (2.9)$$

dient als Zugehörigkeitsfunktion für eine Partition  $l$ , wobei  $z_{l,i}$  mit  $i = 1, 2, \dots, M$  die Koordinaten des Zentrums der Partition beschreiben und die Standardabweichungen  $\sigma_{l,i}$  proportional zur Ausdehnung der Partition in der Dimension  $i$  gewählt werden. Um die lokale Interpretierbarkeit des LMN sicherzustellen und den Beitrag jedes lokalen Modells zu einer Modellvorhersage bewerten zu können, wird für die

Gültigkeitsfunktionen  $\sum_{l=1}^L \Phi_l(\underline{u}) = 1 \forall \underline{u}$  gefordert. Diese Forderung wird durch die Normierung der Zugehörigkeitsfunktion erfüllt und führt zu den Gültigkeitsfunktionen:

$$\Phi_l(\underline{u}) = \frac{\mu_l(\underline{u})}{\sum_{j=1}^L \mu_j(\underline{u})}. \quad (2.10)$$

Als lokale Modelle werden üblicherweise affine Modelle

$$g_l(\underline{u}) = b_{l,0} + b_{l,1} \cdot u_1 + \dots + b_{l,M} \cdot u_M \quad (2.11)$$

verwendet. Diese stellen mit ihren  $M + 1$  Parametern einen guten Kompromiss zwischen der Anzahl an Parametern pro lokalem Modell und der Anzahl an benötigten lokalen Modellen dar, die zusammen die Komplexität des Gesamtmodells bestimmen [46]. Bild 2.6 stellt für zwei Dimensionen beispielhaft die zuvor beschriebenen Elemente eines LMN dar.

Werden lokale Modellstrukturen wie das in Gleichung (2.11) beschriebene Modell verwendet, die linear in den Parametern sind, ist eine separate Schätzung der Modellparameter mittels gewichteter Methode der kleinsten Fehlerquadrate möglich [80].

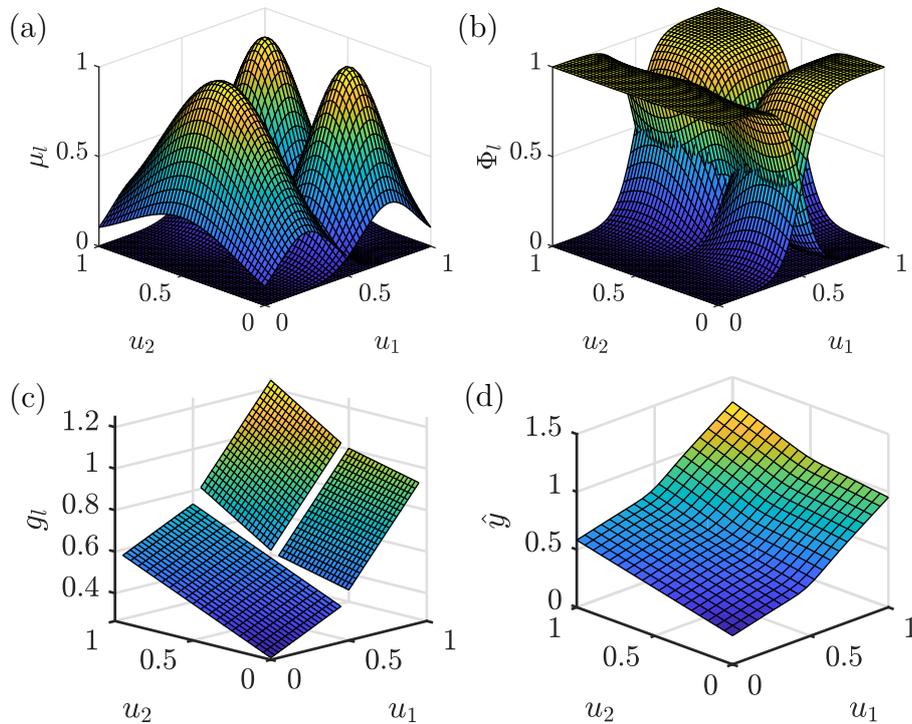


Bild 2.6: Beispielhafte Darstellung eines LMN für zwei Dimensionen mit (a) den Zugehörigkeitsfunktionen  $\mu_l$ , (b) den Gültigkeitsfunktionen  $\Phi_l$ , (c) den lokalen Modellen  $g_l$  und (d) dem Modellausgang  $\hat{y}$  (in Anlehnung an [80]).

Hierbei wird die Gültigkeitsfunktion für jede Partition  $l$  an allen Datenpunkten ausgewertet und in Form einer Gewichtungsmatrix

$$\underline{Q}_l = \text{diag}([\Phi_l(\underline{u}(1)) \ \Phi_l(\underline{u}(2)) \ \cdots \ \Phi_l(\underline{u}(N))]) \quad (2.12)$$

zusammengefasst. Die Modellparameter eines jeden lokalen Modells  $g_l(\underline{u})$  ergeben sich mit dieser Gewichtungsmatrix als Erweiterung der Lösung der klassischen Methode der kleinsten Quadrate (vgl. Gleichung (2.7)) zu:

$$\hat{\underline{\theta}}_l = [b_{l,0} \ b_{l,1} \ \cdots \ b_{l,M}]^T = (\underline{X}^T \underline{Q}_l \underline{X})^{-1} \underline{X}^T \underline{Q}_l \underline{y}. \quad (2.13)$$

Die Regressionsmatrix  $\underline{X}$  und der Ausgangsgrößenvektor  $\underline{y}$  setzen sich analog zu Gleichung (2.6) aus allen Trainingsdatenpunkten zusammen.

LOLIMOT ist trotz der einfachen lokalen Modellstrukturen in der Lage beliebige glatte nichtlineare Zusammenhänge abzubilden. So führt das verwendete Partitionierungsverfahren dazu, dass nichtlineare Bereiche im Eingangsraum solange unterteilt werden, bis insgesamt eine hohe Modellgüte erreicht ist. Durch die Verwendung einer einfachen Heuristik zur Partitionierung und die Schätzung der lokalen Modellparameter mittels gewichteter Methode der kleinsten Quadrate wird insgesamt ein geringer Rechenaufwand erzielt [80].

### 2.2.3 Gaußprozessmodelle

Eine Modellklasse, die eine noch höhere Flexibilität als lokale Modellnetze bietet, sind *Gaußprozessmodelle* (GPM). Diese zählen zu den Bayes'schen Methoden und spiegeln mit bedingten Wahrscheinlichkeiten mögliche funktionale Zusammenhänge wider. Unter Einbeziehung einer *a-priori*-Wahrscheinlichkeitsdichte als Vorwissen, wird das für die gegebenen Daten wahrscheinlichste Modell bestimmt. GPM zählen zu den nichtparametrischen Modellen und erfordern keine Festlegung einer Modellstruktur. Stattdessen besitzt ein GPM so viele Parameter wie Datenpunkte, wodurch mehr Datenpunkten auch eine höhere Modellflexibilität ermöglichen [88]. Nachfolgend wird die Anwendung von GPM zur Regression beschrieben.

Ein grundlegendes Element von GPM ist die Verwendung von *Kernel*-Funktionen. Auf jedem Datenpunkt  $\underline{u}(k)$  mit  $k = 1, 2, \dots, N$  im Eingangsraum wird eine Kernel-Funktion  $K(\underline{u}(k), \underline{u}(l))$  platziert, die zur Messung der Ähnlichkeit des jeweiligen Datenpunkts  $\underline{u}(k)$  zu allen anderen Datenpunkten  $\underline{u}(l)$  mit  $l = 1, 2, \dots, N$  dient.

Hierbei liegt die Annahme zugrunde, dass aus einer großen Ähnlichkeit zwischen zwei Punkten im Eingangsraum auch eine große Ähnlichkeit der zugehörigen Ausgangsgrößen  $y(k)$  und  $y(l)$  resultieren soll. Dies kann in Form einer Kovarianz der Ausgangsgrößen  $\text{cov}\{y(k), y(l)\} = K(\underline{u}(k), \underline{u}(l))$  formuliert werden [80]. Alle Kovarianzen zwischen jeweils zwei Datenpunkten im Trainingsdatensatz können in Form einer Kovarianzmatrix

$$\underline{K} = \begin{bmatrix} K(\underline{u}(1), \underline{u}(1)) & K(\underline{u}(1), \underline{u}(2)) & \cdots & K(\underline{u}(1), \underline{u}(N)) \\ K(\underline{u}(2), \underline{u}(1)) & K(\underline{u}(2), \underline{u}(2)) & \cdots & K(\underline{u}(2), \underline{u}(N)) \\ \vdots & \vdots & \ddots & \vdots \\ K(\underline{u}(N), \underline{u}(1)) & K(\underline{u}(N), \underline{u}(2)) & \cdots & K(\underline{u}(N), \underline{u}(N)) \end{bmatrix} \quad (2.14)$$

zusammengefasst werden. Wird zusätzlich weißes Rauschen  $n$  am Prozessausgang berücksichtigt, beeinflusst dies die Diagonalelemente der Kovarianzmatrix

$$\underline{K} \rightarrow \underline{K} + \hat{\sigma}_n^2 \underline{I} \quad (2.15)$$

zu denen die angenommene Rauschvarianz  $\hat{\sigma}_n^2$  addiert wird. Mit dieser Kovarianzmatrix ergibt sich die a-priori-Wahrscheinlichkeitsdichte für den Modellausgang

$$\hat{\underline{y}} = [\hat{y}(1) \ \hat{y}(2) \ \cdots \ \hat{y}(N)]^T \sim \mathcal{N}(\underline{0}, \underline{K}), \quad (2.16)$$

wobei hier eine multivariate Normalverteilung mit dem Mittelwert null angenommen wird. Diese a-priori-Wahrscheinlichkeitsdichte beinhaltet Vorwissen über das Modellverhalten, wie zum Beispiel die Glattheit des Modellausgangs. Dieses Vorwissen wird über die Wahl der Kernel-Funktion und deren Kernel-Parameter eingebracht. Neben stationären Kernel-Funktionen, die identisches Verhalten im gesamten Eingangsraum aufweisen, können auch nicht stationäre Kernel-Funktionen verwendet werden, deren Charakteristik von der Position im Eingangsraum abhängt. Wird z. B. ein periodisches Prozessverhalten erwartet, kann dies mit einer periodischen Kernel-Funktion abgebildet werden [88]. Für die meisten Modellierungsaufgaben sind nicht periodische, stationäre Kernel-Funktion, wie ein Gaußkern

$$K(\underline{u}(k), \underline{u}(l)) = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{i=1}^M \frac{(u_i(k) - u_i(l))^2}{\sigma_i^2}\right), \quad (2.17)$$

üblich. Dieser wird über  $M + 1$  Kernel-Parameter beeinflusst, bei denen es sich um die *Signal-Standardabweichung*  $\sigma_f$  und individuelle *Kernel-Standardabweichungen*  $\sigma_i$  für jede der  $M$  Eingangsgrößen handelt.

Soll mit dem GPM der Ausgangswert  $y^*$  für einen Punkt im Eingangsraum  $\underline{u}^*$  vorhergesagt werden, wird die multivariate Normalverteilung aus Gleichung (2.16) zu

$$\begin{bmatrix} \hat{y} \\ \hat{y}^* \end{bmatrix} \sim \mathcal{N}(\underline{0}, \tilde{K}), \text{ mit } \tilde{K} = \begin{bmatrix} \underline{K} & \underline{k}(\underline{u}^*) \\ \underline{k}^T(\underline{u}^*) & K(\underline{u}^*, \underline{u}^*) \end{bmatrix} \quad (2.18)$$

erweitert. Alle Kovarianzen zwischen dem vorherzusagenden Punkt und den Datenpunkten im Trainingsdatensatz sind hierbei in einem Kovarianzvektor  $\underline{k}(\underline{u}^*) = [K(\underline{u}^*, \underline{u}(1)) \ K(\underline{u}^*, \underline{u}(2)) \ \dots \ K(\underline{u}^*, \underline{u}(N))]^T$  zusammengefasst. Mit den für alle Trainingsdatenpunkte gemessenen Ausgangswerte  $\underline{y}$  kann die multivariate a-priori-Wahrscheinlichkeitsdichte auf eine univariate Wahrscheinlichkeitsdichte für den vorherzusagenden Ausgangswert

$$\hat{y}^* \sim \mathcal{N}\left(\underline{k}^T(\underline{u}^*) \underbrace{\underline{K}^{-1} \underline{y}}_{\underline{\theta}}, K(\underline{u}^*, \underline{u}^*) - \underline{k}^T(\underline{u}^*) \underline{K}^{-1} \underline{k}(\underline{u}^*)\right) \quad (2.19)$$

reduziert werden. Diese wird als *a-posteriori*-Wahrscheinlichkeitsdichte bezeichnet. Die Vorhersage einer Wahrscheinlichkeitsdichte ermöglicht es, die Unsicherheit der Vorhersage zu beurteilen und beispielsweise in Form eines 95 %-Konfidenzintervalls anzugeben. Ein GPM kann alternativ auch als parametrisches Modell mit dem Parametervektor  $\underline{\theta} = \underline{K}^{-1} \underline{y}$  beschrieben werden, das dieselbe Anzahl an Parametern  $N$  wie Trainingsdatenpunkte besitzt. Der Mittelwert der vorhergesagten Wahrscheinlichkeitsdichte  $\underline{k}^T(\underline{u}^*) \underline{\theta}$  kann so als eine Überlagerung der Kernels aus dem Kovarianzvektor  $\underline{k}(\underline{u}^*)$  gewichtet mit den Parametern  $\underline{\theta}$  interpretiert werden [80].

Das Modellverhalten eines GPM wird in starkem Maße von dessen Hyperparametern beeinflusst. Wird ein Gaußkern gemäß Gleichung (2.17) verwendet, handelt es sich hierbei um die  $M + 1$  Kernel-Parameter  $\sigma_f$  und  $\sigma_i$  mit  $i = 1, 2, \dots, M$ , sowie die geschätzte Standardabweichung des Rauschens  $\hat{\sigma}_n$ . Diese Parameter müssen während des Modelltrainings bestimmt werden, wozu die *Log-Marginal-Likelihood*-Funktion

$$L = - \underbrace{\frac{1}{2} \underline{y}^T (\underline{K} + \hat{\sigma}_n^2 \underline{I})^{-1} \underline{y}}_{\text{Anpassung an Trainingsdaten}} - \underbrace{\frac{1}{2} \ln(\det(\underline{K} + \hat{\sigma}_n^2 \underline{I}))}_{\text{Strafe für Komplexität}} - \underbrace{\frac{N}{2} \ln(2\pi)}_{\text{Konstante}} \quad (2.20)$$

maximiert wird. Die drei Terme in dieser Funktion können als Maß für die Anpassung des Modells an die Trainingsdaten, ein Strafterm für die Komplexität des Modells und eine Konstante interpretiert werden. Durch die Optimierung der Hyperparameter ergibt sich ein Kompromiss, wobei weniger komplexe Modelle bevorzugt werden. Zusätzlich übt die geschätzte Rauschvarianz  $\hat{\sigma}_n^2$  einen regularisierenden Effekt auf das

Modell aus, der eine Überanpassung verhindert. Es ergibt sich eine effektive Anzahl an Parametern kleiner  $N$ , die die Flexibilität des Modells einschränkt [80]. Große Werte von  $\hat{\sigma}_n^2$  bewirken eine starke Regularisierung und führen somit zu einfachen Modellen.

Bild 2.7 stellt beispielhaft den Einfluss der Hyperparameter auf das Modellverhalten dar. Während das Modell durch einen kleinen Wert von  $\hat{\sigma}_n$  in (a) und (b) annähernd interpoliert, sind in (c) und (d) Abweichungen zu den Trainingsdatenpunkten durch die Regularisierung sichtbar. Kleine Werte der Kernel-Standardabweichung  $\sigma_1$  in (a) und (c) ermöglichen schnelle Änderungen der Funktionswerte mit ansteigendem  $u_1$ , während große Werte von  $\sigma_1$  in (b) und (d) nur langsame Änderungen zulassen.

## 2.3 Beurteilung der Eingangsraumabdeckung

Bei der Modellierung realer Prozesse stellt sich die Frage, ob zuvor erfasste Datensätze für die Modellerstellung geeignet sind, oder gesonderte Versuche notwendig sind.

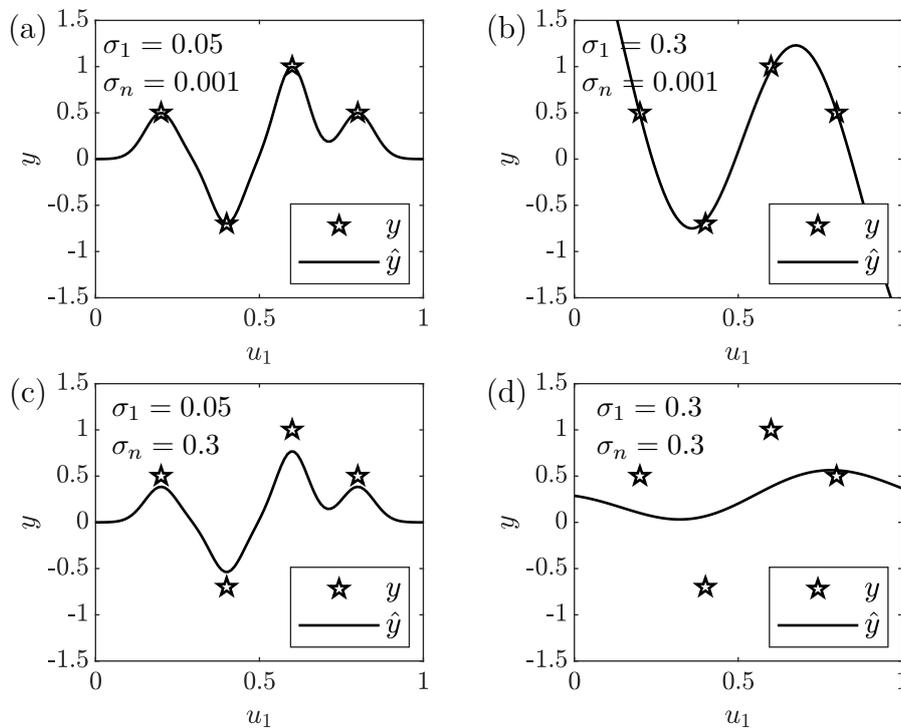


Bild 2.7: Vorhersagen eines GPM für eine Eingangsgröße  $u_1$  mit  $\sigma_f = 1$ , verschiedenen Kernel-Standardabweichungen (links:  $\sigma_1 = 0.05$ , rechts:  $\sigma_1 = 0.3$ ) und Standardabweichungen des Rauschens (oben:  $\sigma_n = 0.001$ , unten:  $\sigma_n = 0.3$ ).

Zur Beantwortung dieser Frage kann die Abdeckung des Eingangsraums mit Datenpunkten analysiert werden. Selbst wenn umfangreiche Datensätze vorhanden sind, können sich die Datenpunkte auf kleine Bereiche des Eingangsraums konzentrieren und Cluster bilden, woraus ungenaue Prozessmodelle resultieren. Eine solche konzentrierte Datenverteilung ist typisch für industrielle Prozesse, die in festen Arbeitspunkten betrieben werden. Bild 2.8 stellt in (a) exemplarisch eine Datenverteilung mit Clustern im Eingangsraum und in (b) eine Gleichverteilung mit derselben Anzahl an Datenpunkten gegenüber. In (c) und (d) sind die Modellausgänge von GPM dargestellt, deren Trainingsdaten mit diesen Datenverteilungen erzeugt wurden. Das Prozessverhalten entspricht in diesem Beispiel einer zweidimensionalen Gaußfunktion. Neben Abweichungen des Modellausgangs ist in (c) insbesondere die Aufweitung

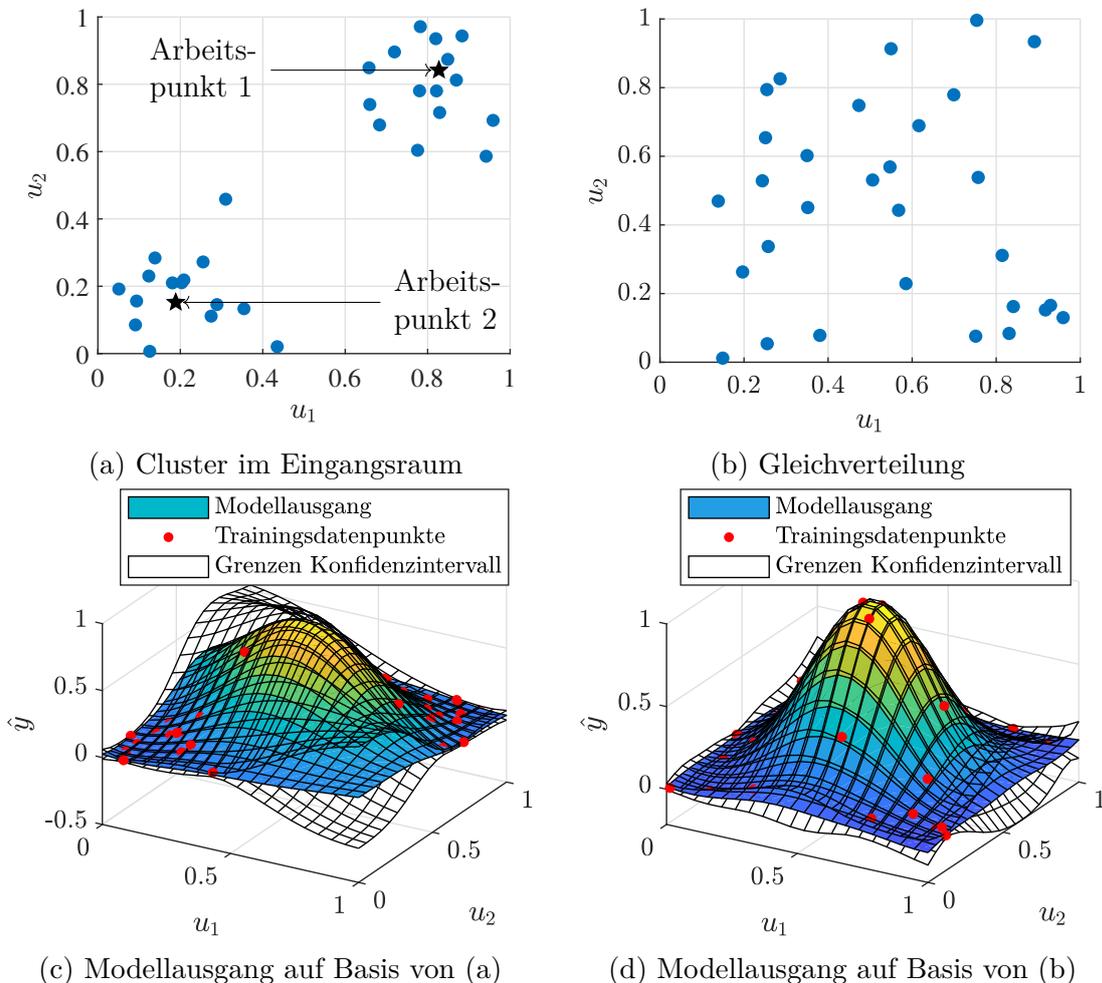


Bild 2.8: Gegenüberstellung der Eingangsdatenverteilungen (a) und (b) mit  $N = 30$  Punkten und den resultierenden Modellausgängen (c) und (d) von GPM mit einer zweidimensionalen Gaußfunktion als zu modellierenden Prozess.

des 95 %-Konfidenzintervalls in großen Bereichen des Eingangsraums sichtbar, die eine größere Modellunsicherheit durch die verwendete Datenverteilung zeigt.

Zur Erstellung von datenbasierten Modellen sind je nach Prozess und gewählter Modellklasse unterschiedliche Verteilungen der Trainingsdaten vorteilhaft. In dieser Arbeit werden Datenverteilungen betrachtet, die möglichst große Bereiche des Eingangsraums gleichmäßig mit Datenpunkten abdecken. Solche Datenverteilungen sind für verschiedene Modellklassen geeignet, wobei insbesondere flexible Modelle wie GPM von einer gleichmäßigen Datenverteilung profitieren [87]. Zur Bewertung der Gleichförmigkeit von Datensätzen ist die *Diskrepanz* ein übliches Kriterium, das die Punktdichte in verschiedenen Bereichen des Eingangsraums vergleicht. Es existieren zahlreiche Ansätze zur Berechnung einer Diskrepanz, für die auf [50] verwiesen sei. Nachfolgend werden weitere Kriterien zur Beurteilung der Eingangsraumabdeckung von Datensätzen beschrieben, die in dieser Arbeit verwendet werden.

### 2.3.1 Distanzen zwischen den Versuchspunkten

Ein häufiger Ansatz zur Beurteilung der Raumabdeckung von Versuchsplänen ist die Betrachtung der Distanzen zwischen Versuchspunkten. Je größer die Distanzen jedes Punkts zu seinen Nachbarpunkten sind, desto gleichmäßiger sind die Datenpunkte über den Eingangsraum verteilt. In [67] werden Mittelwerte, Standardabweichung und Extremwerte der Nächsten-Nachbarn-Distanzen herangezogen, um raumfüllende Versuchspläne zu vergleichen. Mit der kleinsten Nächsten-Nachbarn-Distanz

$$d_{\min} = \min_{\underline{u}(k), \underline{u}(l) \in \mathcal{D}, k \neq l} d(\underline{u}(k), \underline{u}(l)) \quad (2.21)$$

in einem Datensatz  $\mathcal{D} = \{\underline{u}(1), \underline{u}(2), \dots, \underline{u}(N)\}$  mit der euklidischen Distanz  $d(\cdot)$  ergibt sich ein Kriterium, das insbesondere zur Optimierung von raumfüllenden Versuchsplänen geeignet ist (siehe Abschnitt 2.4.2). Verschiedene Datensätze, die anhand dieses Kriteriums optimiert wurden, können mittels  $d_{\min}$  hinsichtlich ihrer Gleichförmigkeit verglichen werden. Zur Beurteilung von Datenverteilungen, die durch andere Mechanismen erzeugt werden, ist dieses Kriterium allerdings nicht geeignet, da es ausschließlich von einem Punktpaar abhängt. So wird die Gleichförmigkeit eines raumfüllenden Versuchsplans als gering bewertet, wenn nur ein einzelner Punkt nah an einem anderen Punkt liegt [80]. Ein solcher Fall mit nur einem ungünstig eingefügten Punkt (orange) ist in Bild 2.9 dargestellt. Die gleichmäßige Abdeckung des Eingangsraums durch die blauen Punkte kann durch  $d_{\min}$  nicht detektiert werden.

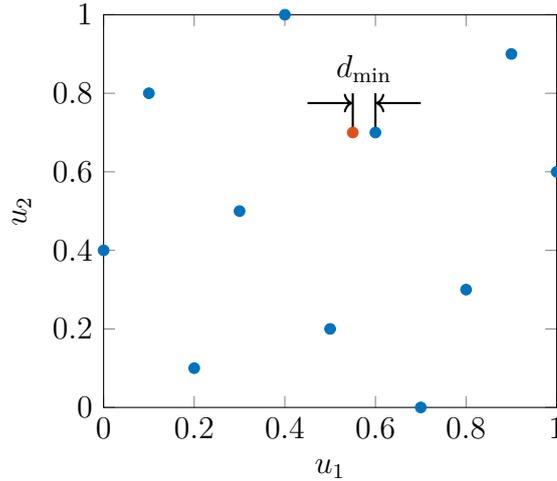


Bild 2.9: Bewertung eines raumfüllenden Versuchsplans (blau) mit dem Kriterium  $d_{\min}$ . Ein zusätzlicher Punkt (orange) verändert  $d_{\min}$  stark.

### 2.3.2 Nächste-Nachbarn-Distanz eines raumfüllenden Versuchsplans

Eine direkte Bewertung von Datenlücken kann prinzipiell durch eine Unterteilung des Eingangsraums in verschiedene Bereiche ermöglicht werden. In [41] werden dazu Voronoi-Zellen auf Basis des gegebenen Datensatzes erstellt und deren räumliche Ausdehnung in verschiedene Bewertungsmaße überführt. Eine Unterteilung des Raums kann auch mit geometrischen Mustern wie beispielsweise (Hyper-)Würfeln erfolgen. Befindet sich kein Punkt des Datensatzes in einem solchen Würfel bzw. mehreren zusammenhängenden Würfeln, liegt eine Lücke im Eingangsraum vor. Bei beiden Verfahren ist eine Auswertung, ob Punkte innerhalb dieser Grenzen liegen, rechenaufwändig und der Aufwand steigt exponentiell mit der Dimension. Daher wird nachfolgend ein Verfahren verwendet, das ausschließlich auf Punktdistanzen basiert und so eine schnelle Auswertung ermöglicht. Die Grundidee dieses Verfahrens ist es, eine hohe Anzahl an Datenpunkten zu generieren, die den Eingangsraum gleichmäßig abdecken und den größten Abstand dieser Punkte zu dem zu untersuchenden Datensatz auszuwerten. Hierzu wird die maximale Nächsten-Nachbarn-Distanz

$$d_S = \max_{\underline{u}_S \in \mathcal{S}} \left( \min_{\underline{u}_D \in \mathcal{D}} d(\underline{u}_S, \underline{u}_D) \right) \quad (2.22)$$

zwischen den Punkten eines raumfüllenden Versuchsplans  $\mathcal{S}$  und dem betrachteten Datensatz  $\mathcal{D}$  als Kriterium verwendet. Der raumfüllende Versuchsplan wird mittels einer Sobol-Sequenz [106] generiert. Diese ermöglicht eine einfache Konstruktion von

Versuchsplänen mit vielen Datenpunkten  $N_S \gg N$  und erzielt eine gleichmäßige Raumabdeckung. Aus der hohen Anzahl an verwendeten Sobol-Punkten resultiert mit  $d_S$  ein repräsentatives Maß für die größte Datenlücke im Eingangsraum. Ein Beispiel für die Bewertung eines Datensatzes mit diesem Kriterium ist in Bild 2.10 gegeben, wobei aus Darstellungsgründen eine geringe Zahl von  $N_S = 64$  Sobol-Punkten (+) verwendet wird. Der orange markierte Sobol-Punkt beschreibt die größte Lücke im Datensatz, woraus sich der Wert von  $d_S$  ergibt.

### 2.3.3 Kerndichteschätzung und Kullback-Leibler-Divergenz

Das nachfolgend beschriebene Kriterium zur Beurteilung der Gleichförmigkeit eines Datensatzes orientiert sich an [56], in dem ein solches Kriterium zur Optimierung von Latin Hypercubes verwendet wird. Es verwendet die *Kullback-Leibler-Divergenz* (KLD) zum Vergleich der diskreten Wahrscheinlichkeitsverteilung eines Datensatzes  $P_D(\underline{u})$  mit der einer Gleichverteilung  $P_G(\underline{u}) = \text{const.}$  Die KLD

$$D_{\text{KL}}(P_D||P_G) = \sum_{\underline{u}} P_D(\underline{u}) \log \frac{P_D(\underline{u})}{P_G(\underline{u})} \quad (2.23)$$

vergleicht  $P_D(\underline{u})$  und  $P_G(\underline{u})$  an allen diskreten Stellen  $\underline{u}$  miteinander und dient somit als Ähnlichkeitsmaß [20]. Kleine Werte zeigen eine große Ähnlichkeit der Verteilungen an, was eine hohe Gleichförmigkeit des zu bewertenden Datensatzes bedeutet.

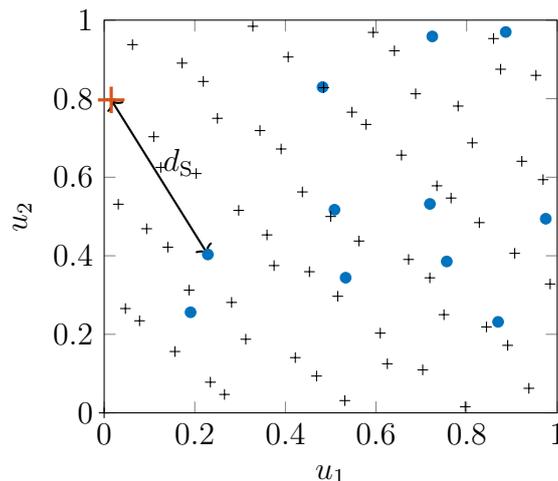


Bild 2.10: Mit den Sobol-Punkten (+) wird die größte Lücke in einem Datensatz detektiert. Der orange markierte Sobol-Punkt weist die größte Nächsten-Nachbar-Distanz  $d_S$  zu einem Punkt im Datensatz auf.

### Kerndichteschätzung

Da die Wahrscheinlichkeitsverteilung eines gegebenen Datensatzes  $\mathcal{D}$  in vielen praktischen Anwendungen unbekannt ist, wird in einem ersten Schritt eine Wahrscheinlichkeitsdichtefunktion  $\hat{p}_{\mathcal{D}}(\underline{u})$  mithilfe einer Kerndichteschätzung aus den Datenpunkten  $\underline{u}(k) \in \mathcal{D}$  mit  $k = 1, 2, \dots, N$  geschätzt. Die Grundidee dieser Schätzung

$$\hat{p}_{\mathcal{D}}(\underline{u}) = \frac{1}{N} \sum_{k=1}^N \frac{1}{\sqrt{(2\pi)^M |\underline{\Sigma}|}} \exp\left(-\frac{1}{2}((\underline{u} - \underline{u}(k))^T \underline{\Sigma}^{-1}(\underline{u} - \underline{u}(k)))\right) \quad (2.24)$$

ist es, auf jedem der  $N$  Datenpunkte eine Kernel-Funktion zu platzieren und diese Kernel-Funktionen additiv zu überlagern, wobei hier Gaußkerne verwendet werden. Alle diese Gaußkerne nutzen dieselbe Kovarianzmatrix  $\underline{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_M^2)$  mit einer individuellen Standardabweichung für jede Eingangsgröße. Zur Bestimmung dieser Standardabweichungen  $\sigma_i$  mit  $i = 1, 2, \dots, M$  wird die *Silverman's rule-of-thumb*

$$\sigma_i = \sigma_{\mathcal{D},i} \cdot \left(\frac{4}{M+2}\right)^{\frac{1}{M+4}} \cdot N^{-\frac{1}{M+4}} \quad (2.25)$$

verwendet [104], wobei  $\sigma_{\mathcal{D},i}$  die Standardabweichung der Daten in der Dimension  $i$  bezeichnet. Da der Eingangsraum einen beschränkten Wertebereich für jede Eingangsgröße aufweist, ergeben sich aus den unbeschränkten Kernel-Funktionen Verzerrungen der geschätzten Wahrscheinlichkeitsdichte an den Rändern des Eingangsraums. So weisen die auf den Datenpunkten platzierten Kernel-Funktionen auch außerhalb des Eingangsraums einen Wert größer null auf. Dies führt insbesondere bei Datenverteilungen mit vielen Punkten am Rand des Eingangsraums dazu, dass nennenswerte Anteile der geschätzten Wahrscheinlichkeitsdichte außerhalb des Eingangsraums liegen und somit abgeschnitten werden. Die Folge ist ein Volumen der Wahrscheinlichkeitsdichtefunktion kleiner eins. Eine übliche Methode, um diesen Verlust an Wahrscheinlichkeitsdichte zu kompensieren (*Boundary Correction*), ist die Spiegelung der über die Ränder hinausragenden Teile der Kernel-Funktionen an diesen Rändern [94].

### Auswertung des Kriteriums

Da keine analytische Lösung für den Vergleich der geschätzten kontinuierlichen Wahrscheinlichkeitsverteilung mit einer Gleichverteilung berechenbar ist, wird stattdessen eine Approximation verwendet. Wie in [20] beschrieben wird die KLD an einer großen Anzahl  $N_{\text{MC}} \gg N$  an Punkten ausgewertet werden, die zufällig mittels einer *Monte-*

*Carlo-Methode* erzeugt werden. Um die bei Zufallsverteilungen möglichen Punkthäufungen zu vermeiden und so eine gleichmäßigere Verteilung dieser Punkte zu erzielen, wird eine Sobol-Sequenz als Quasi-Monte-Carlo-Verfahren [15] verwendet. Mit einer Menge von  $N_{MC}$  Monte-Carlo-Punkten  $\mathcal{S}$  ergibt sich für den gegebenen Datensatz die diskrete Wahrscheinlichkeitsverteilung  $P_D(\underline{u}) = \hat{p}_D(\underline{u})/N_{MC}$  mit  $\underline{u} \in \mathcal{S}$  und für die Gleichverteilung  $P_G(\underline{u}) = 1/N_{MC}$ .

Das beschriebene Vorgehen aus Kerndichteschätzung und Auswertung der KLD kann direkt dazu verwendet werden die Gleichförmigkeit von gegebenen Datensätzen zu beurteilen. Eine hohe Gleichförmigkeit ist hierbei für  $D_{KL} \rightarrow 0$  gegeben. Im Gegensatz zur kleinsten Nächsten-Nachbarn-Distanz  $d_{\min}$  wird dieses Maß durch einzelne ungünstig platzierte Datenpunkte nur wenig beeinflusst und ist damit zur Beurteilung beliebiger Datensätze geeignet. Allerdings existiert bei diesem Kriterium eine Einschränkung bezüglich der Dimensionalität des betrachteten Eingangsraums. Bei der Kerndichteschätzung steigt die Anzahl an benötigten Datenpunkten zur Einhaltung eines maximalen Fehlers exponentiell mit der Dimensionalität an. Aus diesem Grund wird eine Verwendung der Kerndichteschätzung bei praktikablen Stichprobengrößen nur bis zu einer maximalen Dimensionalität von fünf empfohlen [95].

### Kombination mit einem Punktselektionsverfahren

Werden Datensätze mit dieser Methodik bewertet, die während des regulären Betriebs industrieller Prozesse aufgenommen wurden, liegen typischerweise sehr große Datenmengen vor, die sich allerdings häufig auf einzelne Arbeitspunkte konzentrieren. Ein künstliches Beispiel eines solchen Datensatzes ist in Bild 2.11a gegeben, in der sich rund die Hälfte der Datenpunkte um einen Arbeitspunkt in der Mitte des Eingangsraums konzentrieren. Diese Punkte tragen kaum zu einem raumfüllenden Versuchsplan bei und besitzen aufgrund der großen Ähnlichkeit zueinander einen geringen Informationsgehalt. Zur Detektion von Punkthäufungen wird in dieser Arbeit die Verwendung des  $D_{KL}$ -Kriteriums in Verbindung mit einem Punktselektionsverfahren vorgeschlagen. Die Punktselektion wählt aus dem gegebenen Datensatz mit  $N$  Punkten die  $N_{Sel} < N$  Punkte aus, die zu der höchsten Gleichförmigkeit führen. Ein solcher reduzierter Datensatz ist beispielhaft für  $N_{Sel} = 40$  in Bild 2.11b zu sehen.

Verschiedene Strategien, um mittels Punktselektion raumfüllende Datensätze zu erzeugen werden in [90] verglichen. Eine dieser Strategien, die einen geringen Rechenaufwand aufweist ist die *Greedy Deletion*. Hierbei wird sequentiell derjenige Punkt

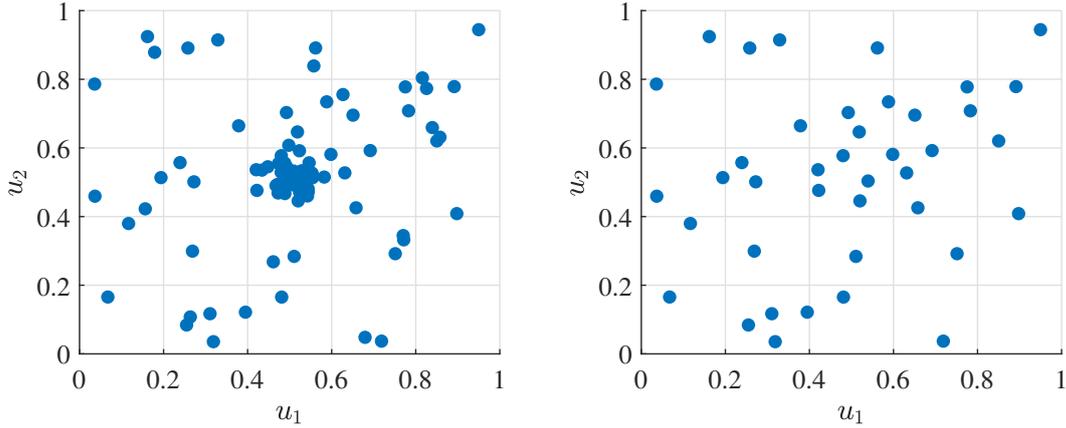
(a) Initialer Datensatz mit  $N = 100$ (b) Reduzierter Datensatz mit  $N_{\text{Sel}} = 40$ 

Bild 2.11: Der für einen industriellen Prozess mit einem Arbeitspunkt beispielhafte Datensatz in (a) wird in (b) mittels Punktselektion reduziert.

entfernt, der den geringsten Beitrag zur gleichmäßigen Abdeckung des Eingangsraums leistet. Die Identifikation dieses Punkts basiert auf dem minimalen euklidischen Abstand (vgl. Gleichung (2.21)). Da mindestens zwei Punkte zueinander den geringsten Abstand aufweisen, kann auf diese Weise kein eindeutiger Punkt erkannt werden. Für jeden Punkt im Datensatz  $\underline{u}(k) \in \mathcal{D}$  wird daher der Abstand zu dessen nächsten Nachbarn

$$d_{\text{NN}}(\underline{u}(k)) = \min_{\underline{u}(l) \in \mathcal{D}_k} (d(\underline{u}(k), \underline{u}(l))) \quad (2.26)$$

bestimmt, wobei  $\mathcal{D}_k$  die Menge aller potentiellen Nachbarpunkte beschreibt und mit  $\mathcal{D}_k = \mathcal{D} \setminus \{\underline{u}(k)\}$  initialisiert wird. Für alle Punkte, bei denen der gleiche minimale Wert von  $d_{\text{NN}}$  vorliegt, wird der nächste Nachbar jeweils aus der Menge aller potentiellen Nachbarpunkte entfernt  $\mathcal{D}_k = \mathcal{D}_k \setminus \{\underline{u}(k)\}$  und Gleichung (2.26) erneut ausgewertet. Auf diese Weise werden die Abstände zu den zweit, dritt usw. nächsten Nachbarn verglichen, bis ein Punkt ein eindeutiges Minimum bei einer dieser Abstände aufweist und somit entfernt werden kann.

Wird die KLD für verschiedene Werte von  $N_{\text{Sel}}$  ausgewertet, kann hieraus die Anzahl an Punkten abgeschätzt werden, die einen wesentlichen Beitrag zur Gleichförmigkeit leisten. Diese Anzahl kann als eine effektive Datenmenge angesehen werden, welche die maßgeblichen Informationen über den Prozess enthält. Die Verläufe von  $D_{\text{KL}}$ ,  $d_s$  und einem kombinierten Kriterium  $D_{\text{KL}} - d_s$  für den Datensatz aus Bild 2.11a sind in Bild 2.12 dargestellt. Dieser Datensatz wird mittels Punktselektion auf bis zu  $N_{\text{Sel}} = 5$  Datenpunkte reduziert. Zur Auswertung der KLD wurden  $N_{\text{MC}} = 2^{14} \gg N$  Punkte verwendet. Das Kriterium  $D_{\text{KL}}$  in (a) ist mit abnehmendem Datensatzumfang

$N_{\text{Sel}}$  annähernd monoton fallend. Hierbei ist im Bereich von 100 bis ca. 50 Punkten eine betragsmäßig große Steigung zu erkennen. Dies deutet auf eine deutliche Verbesserung der Gleichförmigkeit hin und kann durch Auflösung des Clusters in der Mitte des Eingangsraums, in dem ca. 50 Punkte konzentriert sind, erklärt werden. Bei einer weiteren Reduzierung des Datensatzes sind im Vergleich nur geringfügige Änderungen von  $D_{\text{KL}}$  sichtbar, was darauf hindeutet, dass alle verbleibenden Punkte zur Raumabdeckung beitragen. Die Anzahl von 50 Punkten kann für diesen Datensatz als eine effektive Datenmenge identifiziert werden.

Das Kriterium  $d_S$  in (b) spiegelt die größte Lücke im Datensatz wider und wird durch die Punktselektion erst ab  $N_{\text{Sel}} < 13$  Punkten vergrößert. Dies zeigt, dass der gegebene Datensatz stark reduziert werden kann, ohne große Lücken zu erzeugen. Das Kriterium  $D_{\text{KL}}$  wird bei sehr wenigen Datenpunkten ungenau, während das Kriterium  $d_S$  nur bei sehr wenigen Datenpunkten empfindlich ist. Mittels des Kriteriums  $D_{\text{KL}} - d_S$  in (c) lassen sich diese Eigenschaften vorteilhaft kombinieren, wodurch eine obere und eine untere Grenze der Trainingsdatenmengen anhand eines Verlaufs bestimmt werden kann. In Bild 2.12 ist diese Differenz zwischen ca.  $N_{\text{Sel}} = 50$  und ca.  $N_{\text{Sel}} = 15$  Datenpunkten näherungsweise konstant, sodass in diesem Bereich eine gleichförmige Datenverteilung ohne stark anwachsende Lücken im Eingangsraum gegeben ist.

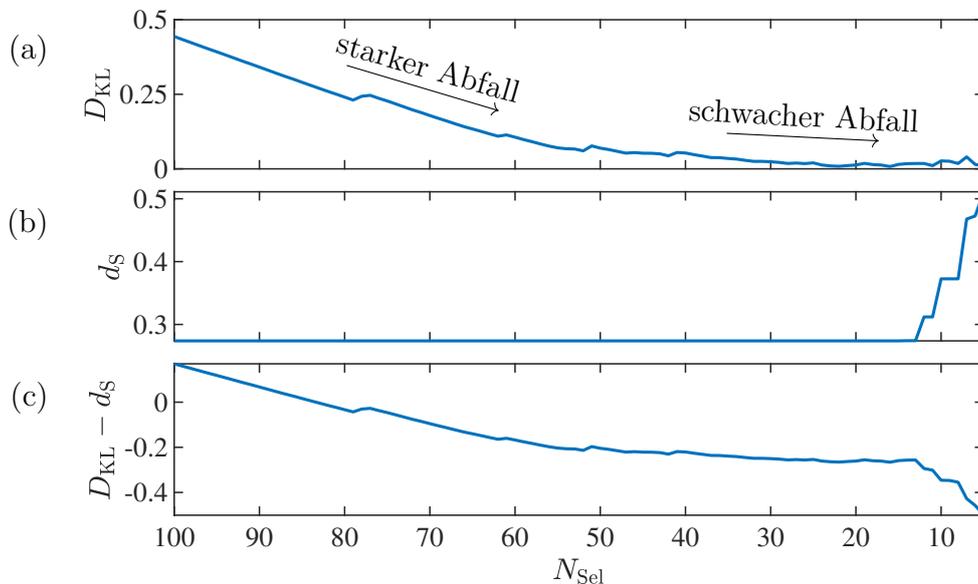


Bild 2.12: Beurteilung des Datensatzes in Bild 2.11 in Verbindung mit der Punktselektion. Die Kriterien (a)  $D_{\text{KL}}$ , (b)  $d_S$  sowie (c) die Kombination  $D_{\text{KL}} - d_S$  sind für verschiedene Anzahlen an selektierten Punkten  $N_{\text{Sel}}$  dargestellt.

## 2.4 Versuchsplanungsmethoden

Die statistische Versuchsplanung (engl. *Design of Experiments*, kurz: DoE) zielt darauf ab, Prozesse systematisch zu vermessen und dabei möglichst relevante Informationen über den Prozess zu erlangen. Auf diese Weise wird mit einem geringen Versuchsaufwand eine hohe Modellgüte ermöglicht. Hierzu werden Versuchspunkte so im Eingangsraum platziert, dass dieser entweder möglichst gleichmäßig mit Datenpunkten abgedeckt ist (*modellfrei*) oder die Datenpunkte so platziert sind, dass die Parameter einer im Vorfeld gewählten Modellstruktur möglichst genau geschätzt werden können (*modellbasiert*). Einen Überblick über verschiedene Verfahren der statistischen Versuchsplanung liefern die Bücher [62, 91, 103]. Durch den Einsatz von Versuchsplanungsverfahren sollen unsichere Modellvorhersagen wie bei dem in Bild 2.8c dargestellten Modell vermieden und eine gleichbleibend hohe Modellgüte im gesamten Eingangsraum ermöglicht werden. Übliche Versuchspläne werden im Vorfeld der Messungen erstellt und sind auf Annahmen bezüglich des Prozessverhaltens angewiesen. Ein alternativer Ansatz sind *sequentielle* Versuchspläne, bei denen weitere Punkte unter Berücksichtigung der Informationen bereits vermessener Versuchspunkte platziert werden. Nachfolgend werden die für diese Arbeit wichtigsten Grundlagen der Versuchsplanungsverfahren erläutert, wobei zwischen modellbasierten, modellfreien und sequentiellen Versuchsplänen unterschieden wird.

### 2.4.1 Modellbasierte Versuchspläne

Modellbasierte Versuchspläne erfordern die Festlegung einer Modellstruktur im Vorfeld der Versuchsplanung. Nachfolgend werden ausschließlich polynomiale Modellstrukturen betrachtet, da diese typischerweise in Kombination mit modellbasierten Versuchsplänen verwendet werden. Die Festlegung des Grads des Polynommodells bzw. der verwendeten Regressoren erfolgt auf Grundlage von Vorwissen oder Annahmen über den zu modellierenden funktionalen Zusammenhang.

#### Klassische Versuchspläne

Klassische Versuchspläne basieren auf geometrischen Mustern und können daher auch als modellfreie Versuchspläne angesehen werden. In der praktischen Anwendung liegen allerdings meist Modellannahmen zugrunde, um einen akzeptablen Versuchsaufwand zu erzielen. Mit diesen Modellannahmen wird eine Auswahl an Effekten

getroffen, die mit einem Versuchsplan erfasst werden können, wobei in *Haupteffekte* und *Wechselwirkungen* der Eingangsgrößen unterschieden wird. Diese Unterscheidung wird am Beispiel des quadratischen Polynommodells

$$\hat{y} = b_0 + \underbrace{b_1 \cdot u_1 + b_2 \cdot u_2 + b_3 \cdot u_1^2 + b_4 \cdot u_2^2}_{\text{Haupteffekte}} + \underbrace{b_5 \cdot u_1 \cdot u_2}_{\text{Wechselwirkung}} \quad (2.27)$$

deutlich. Die Regressoren, die sich aus *einer* Eingangsgröße bzw. deren Potenzen ergeben, werden als Haupteffekte und Kombinationen aus *mehreren* Eingangsgrößen als Wechselwirkungen bezeichnet. Des Weiteren werden für die einzelnen Eingangsgrößen diskrete Werte (*Stufen*) vorgegeben, auf denen die Eingangsgrößen in den Versuchen variiert werden.

Ein *vollfaktorieller Versuchsplan* beinhaltet alle Kombinationen der Stufen für alle Eingangsgrößen. Sind für jede Eingangsgröße nur zwei Stufen definiert, kann dieser Versuchsplan alle linearen Haupteffekte und Wechselwirkungen zwischen jeweils zwei Eingangsgrößen erfassen. Für ein auf Basis dieses Versuchsplans erstelltes Polynommodell sollten dementsprechend nur diese Regressoren bzw. eine Untermenge hiervon ausgewählt werden. *Teilfaktorielle Versuchspläne* ermöglichen es durch eine gezielte Auswahl von relevanten Effekten den Versuchsaufwand gegenüber einem vollfaktoriellen Versuchsplan zu reduzieren. Hierzu werden typischerweise Mehrfachwechselwirkungen zwischen den Eingangsgrößen aufgrund der geringen praktischen Relevanz vernachlässigt [40]. Wird quadratisches Verhalten angenommen, sind bei einem faktoriellen Versuchsplan mindestens drei Stufen pro Eingangsgröße notwendig. Weitere typische Versuchspläne für quadratische Polynommodelle sind *Central-Composite-Designs* mit einem üblicherweise mehrfach vermessenen Zentralpunkt oder *Box-Behnken-Designs*, bei denen Versuchspunkte in den Ecken des Eingangsraums vermieden werden, da diese Bereiche für viele Prozesse kritisch sind [103].

## Optimale Versuchspläne

Optimale Versuchspläne bieten mehr Möglichkeiten an den jeweiligen Anwendungsfall angepasst zu werden, als klassische Versuchspläne. So kann die Anzahl an Versuchspunkten  $N$  und die polynomiale Modellstruktur frei vorgegeben werden. Weitere Vorgaben, wie unterschiedlich viele Stufen bzw. kontinuierliche Werte für die einzelnen Eingangsgrößen oder unzulässige Bereiche im Eingangsraum können auf einfache Weise im Versuchsplan berücksichtigt werden [62]. Optimale Versuchspläne nutzen ein Optimalitätskriterium, um die Versuchspunkte so zu platzieren, dass das

Modell bzw. dessen Parameter möglichst kleine stochastische Fehler aufweisen. Die Optimalitätskriterien beziehen sich daher auf die Parameter eines Modells mit einer vorgegebenen Struktur. Am populärsten sind *D-optimale* Versuchspläne, bei denen das Volumen des Konfidenzbereichs des Parametervektors  $\hat{\theta}$  minimiert wird. Dies entspricht einer Minimierung der Determinante  $|(\underline{X}^T \underline{X})^{-1}|$ . Für weitere optimale Versuchspläne wie A-, G-, oder V-optimale Versuchspläne sei auf [103] verwiesen. Die Anwendung dieser verschiedenen Kriterien erfolgt typischerweise auf Basis einer Kandidatenmenge, die alle realisierbaren Versuchspunkte enthält. Zur Optimierung werden wiederholt Tauschvorgänge zwischen den Versuchspunkten und dieser Kandidatenmenge ausgeführt. Einen Überblick über zu diesem Zweck entwickelte Algorithmen gibt [81].

Die Wahl einer Modellstruktur im Vorfeld der Messungen stellt für die praktische Anwendung dieser Versuchsplanungsverfahren eine große Herausforderung dar. Wird beispielsweise ein Polynommodell mit einem zu niedrigen Grad gewählt, kann dieses Modell die funktionalen Zusammenhänge des Prozesses nur unzureichend abbilden, woraus ein hoher Biasfehler resultiert. Aus diesem Grund wird in [80] der Einsatz von optimalen Versuchsplänen in Verbindung mit Polynommodellen nur bei sehr starkem Prozessrauschen oder bekannter Modellstruktur z. B. auf Grundlage physikalischer Gesetzmäßigkeiten empfohlen.

## 2.4.2 Modellfreie Versuchspläne

Modellfreie Versuchspläne beruhen auf keinen Annahmen über die Modellstruktur. Stattdessen werden Datenverteilungen angestrebt, die den Eingangsraum möglichst gleichförmig abdecken. Diese Art von Versuchsplänen wird als raumfüllend bezeichnet. Raumfüllende Versuchspläne eignen sich insbesondere dann, wenn unzureichendes Vorwissen über die funktionalen Zusammenhänge zwischen Ein- und Ausgangsgrößen vorliegt, sodass keine spezifische Modellstruktur festgelegt werden kann. Sie können genutzt werden, um auf Basis eines Datensatzes Modelle aus verschiedenen Klassen zu evaluieren und sind besonders für flexible Modellklassen wie lokale Modellnetze oder Gaußprozessmodelle vorteilhaft. Ein typisches Anwendungsfeld raumfüllender Versuchspläne sind deterministische Computersimulationen [91]. Sie wurden aber auch erfolgreich zur Modellierung industrieller Prozesse, beispielsweise in der Chemieindustrie eingesetzt [68]. Obwohl keine Modellstruktur zugrunde gelegt wird, können raumfüllende Versuchspläne auf Basis von Vorwissen an einen Prozess angepasst werden. Hierzu können zum Beispiel Grenzen im Eingangsraum

berücksichtigt werden oder Eingangsgrößen unterschiedlich skaliert werden [72]. Zur Erstellung raumfüllender Versuchspläne existieren eine Vielzahl an Verfahren, wie beispielsweise *Uniform Designs*, bei denen die Diskrepanz zwischen der Verteilung der Versuchspunkte und einer Gleichverteilung minimiert wird [57]. Auch *Sphere Packing*-Verfahren, die möglichst viele geometrische Objekte in einem vorgegebenen Raum anordnen [51], können hierzu genutzt werden. In dieser Arbeit werden optimierte *Latin Hypercubes* (LHC) und pseudozufällige Sequenzen zur Erstellung von raumfüllenden Versuchsplänen verwendet.

### Optimierte Latin Hypercubes

Eine häufig verwendete Art raumfüllender Versuchspläne sind optimierte LHC. Ein LHC unterteilt jede Dimension des Eingangsraums in  $N$  Stufen, die meist äquidistant sind. Die Versuchspunkte sind so angeordnet, dass jede Stufe in jeder Dimension nur von jeweils einem Punkt belegt ist. Mit diesem Konstruktionsprinzip erzielen LHC gute Projektionseigenschaften, wenn beispielsweise eine Eingangsgröße bei der Modellbildung als irrelevant betrachtet und nicht im Modell berücksichtigt wird. Ein um beliebig viele Eingangsgrößen reduzierter LHC erfüllt immer noch die LHC-Eigenschaft und wird daher als *non-collapsing* bezeichnet [111]. Zwei verschiedene LHC sind in Bild 2.13 zu sehen, wobei der LHC in (a) den Eingangsraum nur unzureichend abdeckt. Um einen raumfüllenden LHC zu erzeugen, ist eine Optimierung notwendig, die beispielsweise zu dem Versuchsplan in (b) führt. Hierzu sind verschiedene Optimierungskriterien verwendbar, wie das *Minimax*-Kriterium, bei dem die maximale

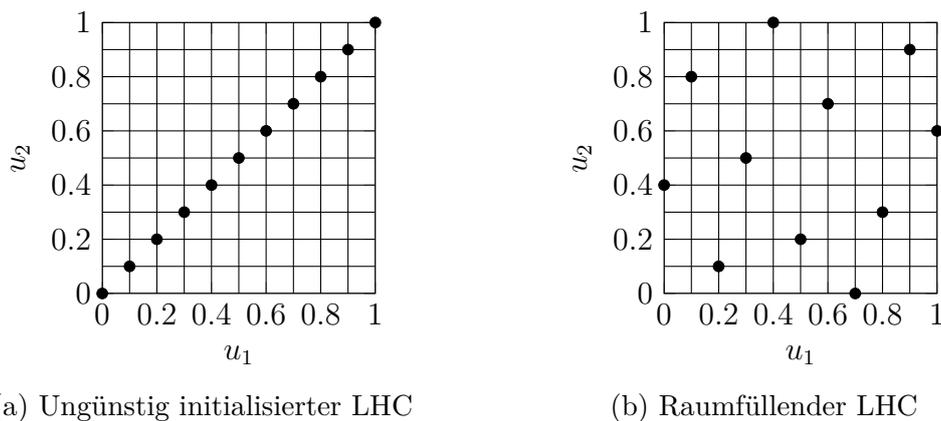


Bild 2.13: Gegenüberstellung von LHC mit  $N = 11$  Punkten im zweidimensionalen Eingangsraum. Jede der  $N$  Stufen in jeder Dimension ist mit exakt einem Punkt belegt. Mittels Optimierung wird in (b) ein raumfüllender LHC erzeugt.

Distanz zwischen den Versuchspunkten minimiert wird, oder das *minimum average reciprocal distance*-Kriterium [91], das auf der mittleren Punktdistanz basiert. Aufgrund des geringen Rechenaufwands wird häufig das *Maximin*-Kriterium

$$\max_{\mathcal{D}}(d_{\min}) = \max_{\mathcal{D}} \left( \min_{\substack{\underline{u}(k), \underline{u}(l) \in \mathcal{D}, \\ k \neq l}} d(\underline{u}(k), \underline{u}(l)) \right) \quad (2.28)$$

bevorzugt, das die kleinste Nächste-Nachbarn-Distanz  $d_{\min}$  (siehe Gleichung (2.21)) im Versuchsplan maximiert.

### EDLS-Algorithmus

Zur Erstellung von raumfüllenden Versuchsplänen wird in dieser Arbeit der *Extended Deterministic Local Search*-Algorithmus (EDLS) [26] verwendet. Hierbei handelt es sich um einen lokalen Suchalgorithmus, der zur Optimierung des Maximinkriteriums wiederholt Koordinatentauschvorgänge in jeweils einer Dimension durchführt. Exemplarisch sind ein solcher Tauschvorgang und die zugehörige Regressionsmatrix für einen zweidimensionalen Eingangsraum in Bild 2.14 dargestellt. Im Beispiel werden die  $u_1$ -Koordinaten der Punkte  $\underline{u}(1)$  und  $\underline{u}(3)$  getauscht, wodurch die minimale Punktdistanz im Versuchsplan von  $d_{\min} = d(\underline{u}(1), \underline{u}(2)) = 0.2$  auf  $d_{\min} = d(\underline{u}(2), \underline{u}(3)) = 0.4$  vergrößert wird. Im EDLS-Algorithmus werden systematisch alle Möglichkeiten zum Koordinatentausch zwischen jeweils zwei Versuchspunkten überprüft. Hierbei werden Tauschvorgänge für die Punkte, die zu der minimalen Punktdistanz im Versuchsplan führen, zuerst betrachtet. Der Algorithmus verfolgt eine *Greedy*-Strategie, sodass eine Verbesserung der minimalen Punktdistanz durch

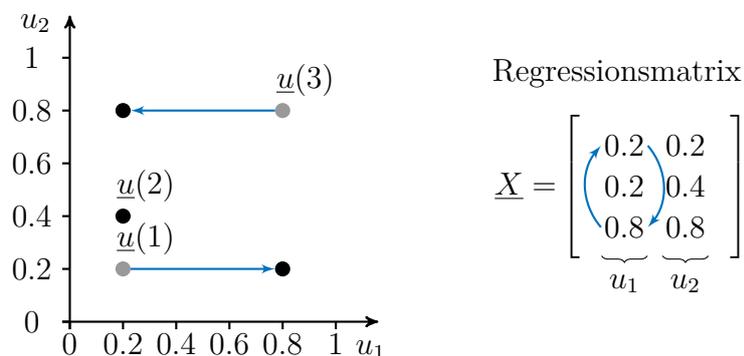


Bild 2.14: Koordinatentausch zweier Punkte  $\underline{u}(1)$  und  $\underline{u}(3)$  in der Dimension  $u_1$ . Durch den Tauschvorgang vergrößert sich die minimale Punktdistanz im Versuchsplan.

einen Tauschvorgang sofort zu einer neuen Iteration des Algorithmus mit dem veränderten Versuchsplan führt. Eine genauere Beschreibung des Algorithmus ist in Anhang A.1 zu finden.

### Pseudozufällige Sequenzen

Werden raumfüllende Versuchspläne mit einer großen Anzahl an Datenpunkten benötigt, ist die Verwendung des zuvor beschriebenen EDLS-Algorithmus aufgrund der hohen Rechenzeit nicht mehr praktikabel. Pseudozufällige Sequenzen können hingegen mit geringem Rechenaufwand erstellt werden und bewirken eine gleichmäßige Abdeckung des Eingangsraums mit Versuchspunkten. Zu den bekanntesten dieser Sequenzen zählen die *Halton*- und *Sobol*-Sequenzen [103]. In dieser Arbeit werden Sobol-Sequenzen verwendet. Diese weisen insbesondere für  $N = 2^g$  Punkte mit  $g \in \mathbb{N}$  gute raumfüllende Eigenschaften auf [106].

### 2.4.3 Sequentielle Versuchspläne

Sequentielle Versuchspläne platzieren Versuchspunkte schrittweise und nutzen dabei Informationen aus vorherigen Schritten. Bei diesen Informationen kann es sich beispielsweise um die ungefähre Position eines Optimums im Eingangsraum handeln. Um eine genauere Bestimmung des Optimums zu ermöglichen wird meist eine höhere Punktdichte in einem Bereich um diese Position angestrebt. Bei einem raumfüllenden Versuchsplan können hierzu beispielsweise mittels des Maximin-Kriteriums zusätzliche Versuchspunkte in diesem Bereich platziert werden, die sowohl untereinander als auch zu den vorhandenen Punkten eine maximale Nächste-Nachbarn-Distanz aufweisen. Ein solches Verfahren ist in [71] beschrieben, wobei die um zusätzliche Punkte zu erweiternde Bereiche manuell festgelegt werden.

Zahlreiche sequentielle Versuchsplanungsverfahren verwenden ein festgelegtes Kriterium, das auf einem Modell basiert. Diese Verfahren werden als *aktives Lernen* (engl. *active learning*) bezeichnet und zielen darauf ab, eine maximale Verbesserung des Modells mit den zusätzlichen Versuchspunkten zu erzielen. Hierbei steht zumeist der Varianzfehler des Modells im Fokus, wie es beispielsweise bei der folgenden Strategie für Gaußprozessmodelle (GPM) der Fall ist. Bei einem GPM kann ein weiterer Versuchspunkt dort platziert werden, wo das Modell das größte Konfidenzintervall aufweist [83], sodass eine möglichst große Reduktion der Modellunsicherheit gegeben

ist. Die Platzierung der Versuchspunkte wird dabei allerdings nur indirekt durch die Optimierung der Hyperparameter des GPM von den gemessenen Ausgangsgrößen beeinflusst, sodass diese Strategie in der Praxis zu einer gleichmäßigen Verteilung der Trainingsdatenpunkte im Eingangsraum führt [80]. Seltener zielen Algorithmen des aktiven Lernens darauf ab, den Biasfehler zu verringern. In [47] wird mit HilotDoE ein solcher Algorithmus mit Fokus auf den Biasfehler beschrieben. Dieser erstellt ein lokales Modellnetz, das in jedem Schritt um ein neues lokales Modell erweitert wird. Die Versuchspunkte werden so platziert, dass alle lokalen Modelle dieselbe Anzahl an Datenpunkten aufweisen und jeder neue Punkt einen maximalen Abstand zu den bereits vermessenen Punkten besitzt.

Aktives Lernen kann neben der Steigerung der Modellgüte auch eingesetzt werden, um unzulässige Bereiche des Eingangsraums zu erkennen. Methoden der *Design Space Exploration* zielen darauf ab, den Eingangsraum zu erkunden, ohne für den Prozessausgang festgelegte Grenzwerte zu unter- oder überschreiten. Ein solches Vorgehen mit zwei parallel erstellten GPM ist in [93] beschrieben. Hierbei wird eine Kombination aus einem Regressionsmodell für die Vorhersage der Ausgangsgröße und einem Klassifikator für die Vorhersage zulässiger Bereiche verwendet.

## 2.5 Zusammenfassung

In diesem Kapitel wurden die Grundlagen der experimentellen Modellbildung vorgestellt. Hierzu wurden Kriterien zur Bewertung und zum Vergleich verschiedener Verfahren beschrieben. Anschließend wurden mit Polynommodellen, lokalen Modellnetzen und Gaußprozessmodellen drei Modellklassen vorgestellt, die eine unterschiedliche Modellflexibilität bieten und sich für viele praktische Modellierungsaufgaben eignen. Sowohl zur Beurteilung gegebener Datensätze als auch zum Vergleich von Versuchsplänen wurden Bewertungsmaße beschrieben, mit denen die Eingangsraumabdeckung charakterisiert werden kann. Abschließend wurden Grundlagen von modellbasierten, modellfreien und sequentiellen Versuchsplänen erläutert, auf denen die in dieser Arbeit beschriebenen Versuchsplanungsmethoden basieren.

## 3 Inkrementelle Modellbildung

Dieses Kapitel beschreibt die *inkrementelle Modellbildung*, die zur Vermessung und Modellierung bestehender mechatronischer Prozesse entwickelt wurde. Dabei werden ein Versuchsplan und ein Prozessmodell in mehreren Schritten aufgebaut, wobei die Anzahl an betrachteten Eingangsgrößen mit jedem Schritt ansteigt. Die geringe Dimensionalität und damit auch Komplexität der Modellierungsaufgabe in den ersten Schritten sowie die Nutzung von Informationen aus vorherigen Schritten sind entscheidende Vorteile dieser Methode und führen zu einer guten praktischen Anwendbarkeit. Das Kapitel basiert auf den Konferenzbeiträgen [112, 113] sowie dem Journal-Beitrag [114].

Zu Beginn wird in Abschnitt 3.1 die Grundidee erläutert. Eine detaillierte Beschreibung der Methodik findet sich in Abschnitt 3.2, wobei verschiedene Modellklassen eingesetzt werden können, die in Abschnitt 3.3 präsentiert werden. In Abschnitt 3.4 werden Vor- und Nachteile der Methodik und der verschiedenen Modellklassen diskutiert. Das Kapitel endet mit einer Zusammenfassung der wichtigsten Erkenntnisse in Abschnitt 3.5. Reale Anwendungen dieser Methodik sind in Abschnitt 5.1 mit der Modellierung einer Heizstrecke und in Abschnitt 5.2 im Rahmen der modellbasierten Optimierung eines Bitumenofens zu finden.

### 3.1 Einführung und Idee

Viele industrielle Prozesse werden in festen Arbeitspunkten betrieben. Dies hat zur Folge, dass der Informationsgehalt der Daten aus dem regulären Betrieb häufig gering ist. Ein anschaulicher Eingangsraum eines möglichen industriellen Prozesses mit zwei Eingangsgrößen ist in Bild 3.1 dargestellt. In unmittelbarer Nähe des Arbeitspunkts ergibt sich ein Betriebsbereich, in dem viele Datenpunkte liegen. Dieser deckt allerdings nur einen Teil des Eingangsraums ab. Modelle, die mit diesen Daten trainiert

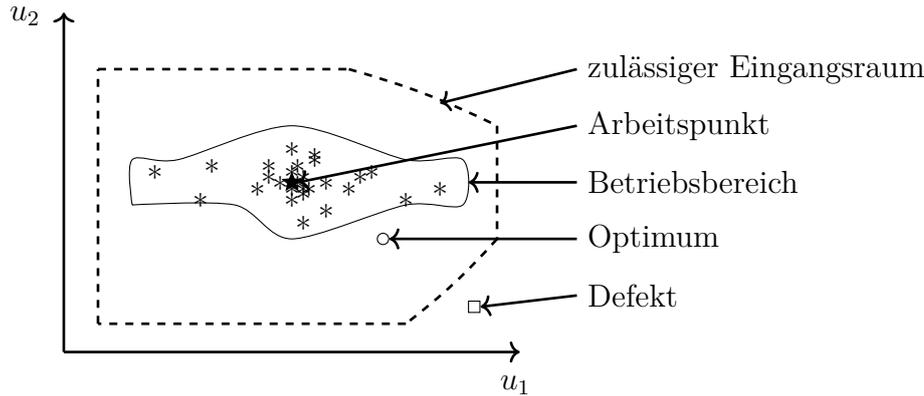


Bild 3.1: Schematische Darstellung des Eingangsraums eines industriellen Prozesses für zwei Eingangsgrößen  $u_1$  und  $u_2$  (in Anlehnung an [112]).

werden, erreichen nur im Betriebsbereich eine hohe Modellgüte, während Modellvorhersagen außerhalb dieses Bereichs durch Extrapolation ungenau sind. Liegt ein Optimum wie in Bild 3.1 außerhalb des Betriebsbereichs, kann dieses mit einem solchen Modell nicht oder nur ungenau bestimmt werden. Um dies zu vermeiden und die Modellqualität zu steigern, müssen folglich Punkte außerhalb des Betriebsbereichs vermessen werden. Zu beachten ist, dass hierbei die Grenzen des zulässigen Eingangsraums, die durch den Prozess vorgegeben sind, nicht überschritten werden dürfen. Wird eine solche Grenze verletzt, kann ein unerwünschtes Prozessverhalten oder sogar ein Defekt an einer Maschine auftreten.

Versuchsplanungsverfahren platzieren Versuchspunkte systematisch, um mit geringem Aufwand viele Informationen über einen Prozess zu erlangen [77]. Typische Versuchspläne, wie sie in Abschnitt 2.4.1 beschrieben sind, werden in einem Schritt erstellt und variieren alle Eingangsgrößen gleichzeitig. Diese gleichzeitige Variation ist zur Erzielung eines hohen Informationsgehalts zu bevorzugen, birgt jedoch die Gefahr, Versuchspunkte außerhalb des zulässigen Eingangsraums zu platzieren. Da die exakten multivariaten Grenzen des zulässigen Eingangsraums zumeist unbekannt sind, sind unzulässige Punkte bei der Versuchsdurchführung schwer zu erkennen. In der Praxis wird daher häufig die einfacher zu überwachende *One-Factor-at-a-Time* (OFaaT) Methode angewandt, bei der nacheinander einzelne Eingangsgrößen variiert werden. Kritische Versuchspunkte in der Nähe von Grenzen im Eingangsraum können so besser erkannt und vermieden werden. OFaaT führt jedoch zu einer ungleichmäßigen Datenverteilung im Eingangsraum [21].

Die aus verschiedenen Verfahren resultierenden Versuchspläne sind in einem zweidimensionalen Eingangsraum in Bild 3.2 gegenübergestellt. Ein Versuchsplan gemäß

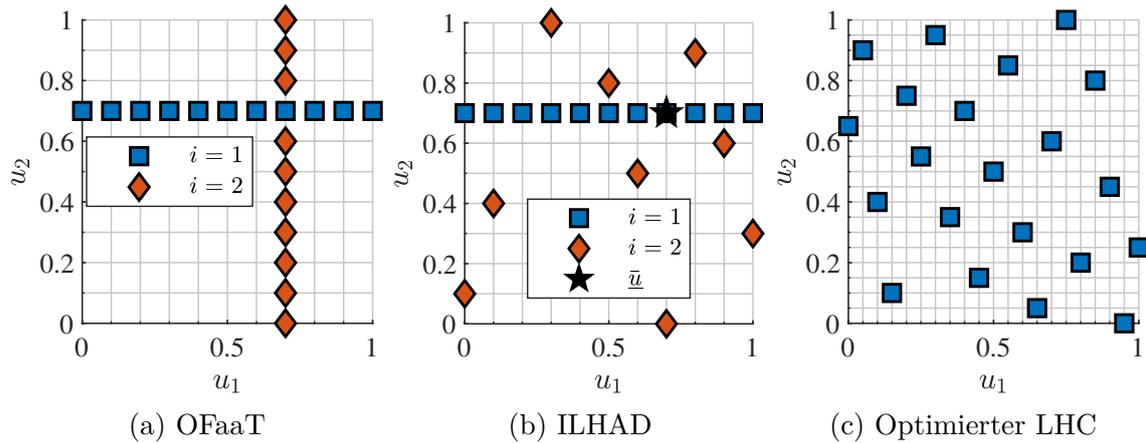


Bild 3.2: Gegenüberstellung drei verschiedener Versuchspläne mit jeweils 21 Versuchspunkten im zweidimensionalen Eingangsraum. Die Schritte  $i = 1$  und  $i = 2$  in (a) und (b), sowie der Arbeitspunkt  $\bar{u}$  in (b) sind dargestellt.

OFaaT wird wie in (a) in mehreren Schritten  $i = 1, 2$  erstellt, wobei jeweils nur eine Eingangsgröße variiert wird. In (c) ist mit einem optimierten *Latin Hypercube* (LHC) ein typischer raumfüllender Versuchsplan dargestellt, der in einem Schritt erstellt wird. Der Versuchsplan in (b) resultiert aus der in diesem Kapitel beschriebenen inkrementellen Modellbildung (engl. *Incremental Latin Hypercube Additive Design*, kurz: ILHAD) und kann als ein Kompromiss in Bezug auf die Abdeckung des Eingangsraums und die Anwendbarkeit in der Praxis zwischen OFaaT und einem raumfüllenden Versuchsplan angesehen werden. Somit ist in Bild 3.2 eine von links nach rechts ansteigende Raumabdeckung zu sehen, die allerdings mit zunehmendem Aufwand bei der Versuchsdurchführung und -überwachung einhergeht.

Ein Versuchsplan gemäß ILHAD wird in mehreren Schritten  $i$  erstellt, wobei eine Untermenge der zur Verfügung stehenden Eingangsgrößen variiert wird. Im Schritt  $i = 1$  können wie in Bild 3.2 identische Versuchspläne mittels ILHAD und OFaaT erzeugt werden. Bei ILHAD wird jedoch in den folgenden Schritten die Anzahl der variierten Eingangsgrößen um eins erhöht und jeweils ein raumfüllender Versuchsplan für die neuen Versuchspunkte erstellt. ILHAD nutzt einen vorgegebenen Arbeitspunkt  $\bar{u}$ , um Werte für alle in einem Schritt noch nicht betrachteten Eingangsgrößen vorzugeben. Ein weiteres Kernelement der inkrementellen Modellbildung ist die schrittweise Erstellung eines Prozessmodells parallel zum Versuchsplan. Dieses Modell wird in jedem Schritt um die Einflüsse der neu betrachteten Eingangsgröße ergänzt, wobei auch Wechselwirkungen mit zuvor betrachteten Eingangsgrößen abgebildet werden. Die einzelnen Elemente der ILHAD-Methodik werden nachfolgend im Detail erläutert.

## 3.2 Methodik

Nachfolgend wird ein Prozess mit insgesamt  $M$  Eingangsgrößen  $u_1, u_2, \dots, u_M$  und einer Ausgangsgröße  $y$  betrachtet, dessen statisches Verhalten mit einem Regressionsmodell abgebildet werden soll. Typische Versuchsplanungsverfahren erfordern eine Auswahl der  $p \leq M$  betrachteten Eingangsgrößen im Vorfeld der Messungen. Die ausgewählten Eingangsgrößen spannen den Eingangsraum auf, für den in diesem Kapitel stets eine Normierung gemäß  $0 \leq u_i \leq 1 \forall i = 1, 2, \dots, p$  angenommen wird. Durch die inkrementelle Einbeziehung weiterer Eingangsgrößen bei ILHAD muss  $p$  nicht im Vorfeld der Versuche festgelegt werden. Dies ermöglicht es, so viele Schritte durchzuführen und damit Eingangsgrößen zu betrachten, bis ein für die Anwendung ausreichendes Modell gegeben ist.

### Rangfolge der Eingangsgrößen

In der Regel haben die Eingangsgrößen einen unterschiedlichen Einfluss auf den Prozessausgang. Daher ist es vorteilhaft, die funktionalen Zusammenhänge der Eingangsgrößen mit den größten Einflüssen zuerst zu modellieren. Unter Umständen kann somit bereits nach wenigen Schritten eine für die Praxis hinreichende Modellgüte erzielt werden und die Modellierung erfolgreich beendet werden. Aus diesem Grund wird im Vorfeld der inkrementellen Modellbildung eine Rangfolge erstellt, in der die Eingangsgrößen gemäß ihrem erwarteten Einfluss auf den Prozessausgang in absteigender Reihenfolge sortiert sind. Diese Rangfolge kann manuell auf der Grundlage von Prozesswissen festgelegt werden, wobei unterstützend auch automatische Verfahren zum Einsatz kommen können. Diese Verfahren basieren z. B. auf Korrelationen oder Informationskriterien und ermöglichen es, die Bedeutung der Eingangsgrößen zu bewerten [42]. Hierfür können Daten aus dem regulären Betrieb verwendet werden, wobei allerdings die typischen Datenverteilungen (vgl. Bild 3.1) die Aussagekraft dieser Verfahren einschränken. Alle Eingangsgrößen werden gemäß der Rangfolge im Eingangsgrößenvektor  $\underline{u} = [u_1 \ u_2 \ \dots \ u_p]^T$  sortiert zusammengefasst, sodass von  $u_1$  die größte Auswirkung auf den Prozessausgang erwartet wird.

### Ablauf der Methodik

Der Ablauf der ILHAD-Methodik ist in Bild 3.3 dargestellt. Nach der initialen Festlegung einer Rangfolge für die Eingangsgrößen werden alle weiteren Aktionen in

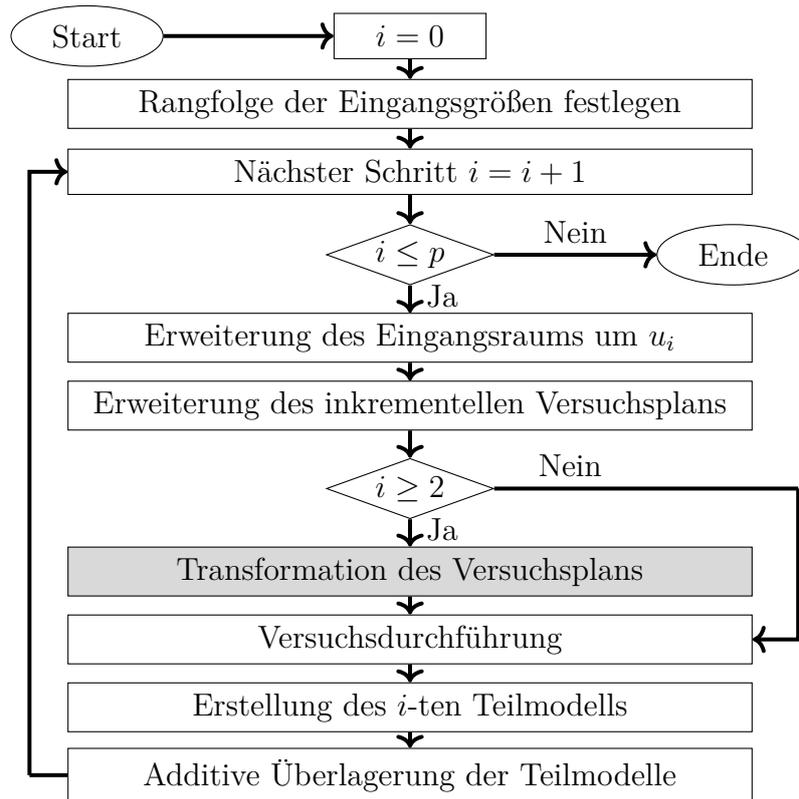


Bild 3.3: Ablauf des Verfahrens zur inkrementellen Modellbildung (in Anlehnung an [113]). Die grau hinterlegte Aktion ist optional.

jedem Versuchsplanungs- und Modellierungsschritt  $i \leq p$  wiederholt durchgeführt. Die Anzahl an durchzuführenden Schritten  $p \leq M$  kann jederzeit z. B. auf Grundlage der verbleibenden Messzeit oder anhand einer geforderten Modellqualität angepasst werden. Zu Beginn jedes Schritts  $i = 1, 2, \dots, p$  wird der betrachtete Eingangsraum um die nächste Eingangsgröße  $u_i$  erweitert. Anschließend wird der inkrementelle Versuchsplan so ergänzt, dass der nun höherdimensionale Eingangsraum möglichst gleichmäßig mit Versuchspunkten abgedeckt wird. Das Konstruktionsprinzip dieses Versuchsplans wird in Abschnitt 3.2.1 erläutert. Die in den Schritten  $i \geq 2$  bereits vorhandenen Modelle können für die weitere Versuchsplanung genutzt werden. Zum Beispiel können Informationen über Nichtlinearitäten bezüglich der zuvor betrachteten Eingangsgrößen aus diesen Modellen extrahiert werden. Diese Informationen können für eine Transformation des Versuchsplans verwendet werden, um die Dichte an Versuchspunkten in stark nichtlinearen Bereichen des Eingangsraums zu steigern. Eine solche Transformation wird in Abschnitt 3.3.2 in Verbindung mit lokalen Modellnetzen beschrieben. Die neuen Versuchspunkte im inkrementellen Versuchsplan werden anschließend am Prozess vermessen, wobei alle noch nicht betrachteten Eingangsgrößen mit  $u_{i+1} = \bar{u}_{i+1}, u_{i+2} = \bar{u}_{i+2}, \dots, u_M = \bar{u}_M$  auf den Werten des Arbeits-

punkts  $\bar{u} = [\bar{u}_1 \ \bar{u}_2 \ \dots \ \bar{u}_M]^T$  konstant gehalten werden. Auf Basis der neuen Versuchsdaten wird anschließend ein Teilmodell erstellt, das additiv zu den bestehenden Teilmodellen hinzugefügt wird (siehe Abschnitt 3.2.2). Hierbei können verschiedene Modellklassen eingesetzt werden, die in Abschnitt 3.3 beschrieben sind.

Die wichtigsten Vorteile, die sich aus dieser Methodik ergeben, sind:

- Durch die schrittweise Erschließung des Eingangsraums und des Prozessverhaltens ist ein intuitives Vorgehen zur Vermessung realer Prozesse gegeben.
- Der Vermessungs- und Modellierungsvorgang kann nach jedem Schritt beendet werden, sobald das Modell alle Anforderungen erfüllt.
- Wissen aus vorherigen Schritten kann zur Unterstützung der Versuchsplanung und der Modellbildung eingesetzt werden.
- In den Versuchen kann sich ein verringerter Stellaufwand ergeben, wenn schwer zu variierende Eingangsgrößen in den ersten Schritten konstant sind.

### 3.2.1 Inkrementelle Versuchsplanung

Das Ziel der Versuchsplanung ist es, in jedem Schritt  $i$  in dem jeweiligen  $i$ -dimensionalen Eingangsraum einen raumfüllenden Versuchsplan zu erzeugen, sodass dieser Raum möglichst gleichmäßig mit Datenpunkten abgedeckt ist. Auf diese Weise wird eine hohe Modellgüte im gesamten betrachteten Eingangsraum angestrebt und die Verwendung flexibler Modellklassen ermöglicht. Die Konstruktion des inkrementellen Versuchsplans erfolgt ausgehend von einem vorgegebenen Arbeitspunkt  $\bar{u}$ , der typisch für den Betrieb des betrachteten Prozesses ist. Zum einen ist so ein Ausgangspunkt gegeben, der sicher im zulässigen Eingangsraum liegt (vgl. Bild 3.1), zum anderen hat der Bereich um einen solchen Arbeitspunkt eine hohe Relevanz für den Betrieb des Prozesses. Industrielle Prozesse werden typischerweise konstruktiv für einen vorgegebenen Arbeitspunkt ausgelegt und weisen in dessen Nähe ein vorgesehenes Verhalten auf. Es ist somit davon auszugehen, dass ein zu bestimmendes Optimum in der Nähe dieses Arbeitspunkts liegt.

#### Konstruktionsprinzip des inkrementellen Versuchsplans

Ein inkrementeller Versuchsplan gliedert sich in mehrere Teilversuchspläne  $\mathcal{D}_i$  mit  $N_i$  Versuchspunkten, die jeweils in einem Schritt  $i$  erstellt werden. Jeder Teilversuchsplan wird als ein zufälliger LHC im  $i$ -dimensionalen Eingangsraum initialisiert.

Zur Erzeugung eines raumfüllenden Versuchsplans wird dieser LHC anschließend optimiert. ILHAD verwendet hierzu den *Extended Deterministic Local Search* (EDLS) Algorithmus [26], der die minimale Distanz zwischen den Versuchspunkten maximiert (siehe Abschnitt 2.4.2). Die zuvor vermessenen Versuchspunkte gehen mit in die Optimierung des neuen LHC ein. Alle im LHC nicht berücksichtigten Eingangsgrößen werden mit  $u_j = \bar{u}_j \forall j = i + 1, i + 2, \dots, M$  auf dem zugehörigen Wert des Arbeitspunkts konstant gehalten.

Ein beispielhafter Versuchsplan im zweiten und dritten Schritt der inkrementellen Modellbildung ist in Bild 3.4 dargestellt. Im Schritt  $i = 1$  ergeben sich für  $\mathcal{D}_1$  durch den LHC äquidistante Punkte (Quadrate) im eindimensionalen Eingangsraum. Dieser eindimensionale Eingangsraum erstreckt sich parallel zur Achse  $u_1$  durch den Arbeitspunkt  $\bar{u}$ . Die LHC-Stufen sind hierbei so gewählt, dass der Arbeitspunkt ein Punkt des LHC ist. Der Versuchsplan in (a) wird im Schritt  $i = 2$  um die Punkte in  $\mathcal{D}_2$  erweitert, die einen optimierten LHC für zwei Dimensionen bilden (Rauten). In (b) wurde der Versuchsplan aus (a) im Schritt  $i = 3$  um die Punkte in  $\mathcal{D}_3$  erweitert (Dreiecke), die einem LHC im dreidimensionalen Raum entsprechen. Dieses Konstruktionsprinzip kann für beliebig viele Dimensionen fortgesetzt werden, sodass in jedem Schritt eine zuvor konstant gehaltene Eingangsgröße durch die Versuchsplanung systematisch variiert wird. Die in (b) grau markierte Ebene mit  $u_3 = \bar{u}_3 = 0.7$  beinhaltet alle Punkte, die in (a) dargestellt sind. Allgemein liegen im Schritt  $i$  alle zuvor vermessenen Versuchspunkte in einem Unterraum des Eingangsraums mit

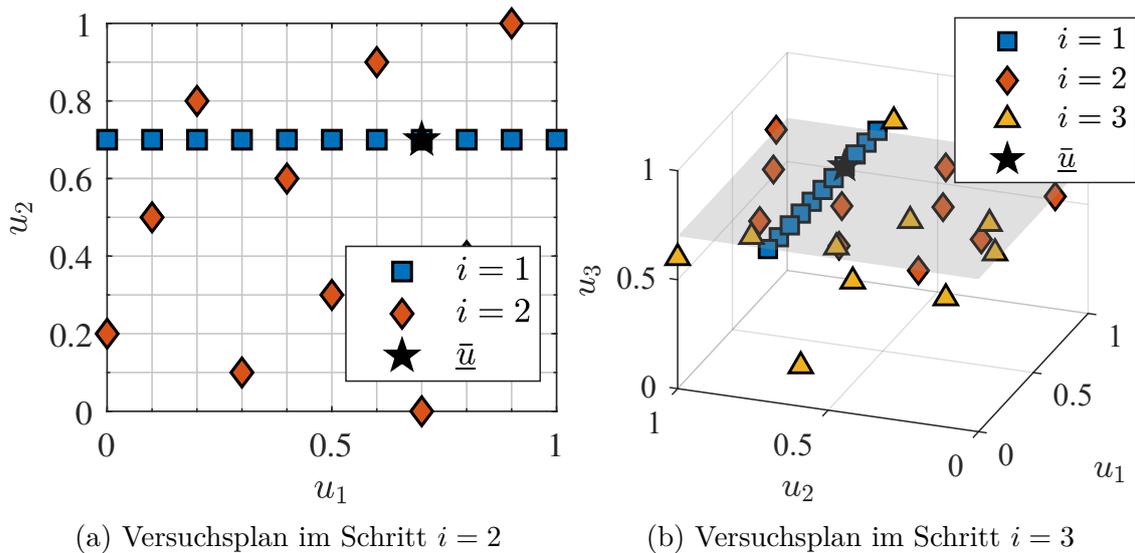


Bild 3.4: Inkrementeller Versuchsplan im (a) zweiten und (b) dritten Schritt mit 11 Stufen für jeden LHC und dem Arbeitspunkt  $\bar{u} = [0.7 \ 0.7 \ 0.7]^T$ . Die graue Ebene in Darstellung in (b) beinhaltet alle Punkte aus (a).

$u_i = \bar{u}_i$ . Werden die LHC-Stufen so gewählt, dass ein Versuchspunkt in diesen Unterraum fällt, kann dieser Punkt aus dem inkrementellen Versuchsplan entfernt werden. Die Vermessung eines solchen Punkts hätte durch die verwendeten Teilmodellstrukturen keinen Einfluss auf das Modell (siehe Abschnitt 3.3). In Bild 3.4 sind daher in den Schritten  $i = 2$  und  $i = 3$  jeweils nur 10 Punkte zu sehen.

### Festlegung der Anzahl an Versuchspunkten

Der lineare Anstieg der Dimensionalität mit jedem Schritt führt zu einem exponentiellen Anstieg des Hypervolumens des Eingangsraums. Würde in jedem Schritt die initiale Dichte der Versuchspunkte beibehalten werden, hätte dies einen exponentiellen Anstieg der Punkte und damit einen für viele Anwendungen nicht praktikablen Versuchsaufwand zur Folge. Dieser Effekt wird als *Fluch der Dimensionalität* bezeichnet [10]. ILHAD bezieht die Rangfolge der Eingangsgrößen mit ein, um den Fluch der Dimensionalität zu vermindern. Je geringer der Einfluss einer Eingangsgröße auf den Prozessausgang ist, desto kleiner ist deren Beitrag zum Modellfehler. In jedem Schritt wird eine Eingangsgröße mit geringerem Einfluss hinzugefügt, sodass die zusätzlichen funktionalen Zusammenhänge ungenauer approximiert werden können, ohne einen übermäßigen Beitrag zum Modellfehler zu leisten. Bei hinreichend großen Unterschieden in der Bedeutung der einzelnen Eingangsgrößen kann folglich eine Verringerung der Punktdichte akzeptiert werden. Die Anzahlen an Versuchspunkten in den einzelnen Schritten des inkrementellen Versuchsplans  $\underline{N} = [N_1 \ N_2 \ \dots \ N_p]^T$  können auf verschiedene Weisen festgelegt werden, die zu einer Verringerung des Versuchsaufwands gegenüber dem exponentiellen Anstieg führen. Beispielsweise kann eine konstante Anzahl oder ein linearer Anstieg mit jedem Schritt vorgegeben werden. Zusätzlich kann Prozesswissen eingesetzt werden, um Schritte mit mehr Versuchspunkten zu versehen, in denen Eingangsgrößen mit stark nichtlinearem Einfluss oder vielen Interaktionen mit anderen Eingangsgrößen einbezogen werden.

### 3.2.2 Inkrementelle Modelle

Die schrittweise Vermessung von Datenpunkten in einem Eingangsraum mit steigender Dimensionalität erfordert eine Modellstruktur, die die zusätzlichen Informationen abbilden kann, ohne dabei zuvor erlangte Informationen zu verlieren. Ähnliche Anforderungen bestehen bei der *Greedy Stagewise Linear Regression* [30], die

zur Lösung von linearen Regressionsproblemen mit vielen Regressoren dient. Hierbei werden Teilmodelle schrittweise erstellt und additiv überlagert. Jedes Teilmodell verwendet eine individuelle Untermenge an Regressoren und korrigiert Fehler des vorherigen Modells. Ein vergleichbares Prinzip wird für ILHAD genutzt und führt zu der nachfolgend beschriebenen Modellstruktur.

### Struktur des Gesamtmodells

Für ILHAD wurde die in Bild 3.5 dargestellte Struktur verwendet, in der mehrere Teilmodelle ein Gesamtprozessmodell bilden. Die einzelnen Teilmodelle  $f_i$  werden nacheinander in jedem der  $i = 1, 2, \dots, p$  Schritte der inkrementellen Modellbildung erstellt und treffen jeweils eine Vorhersage aus einer Untermenge der  $p$  betrachteten Eingangsgrößen. Jedes in einem Schritt  $i$  erstellte Teilmodell  $f_i$  modelliert die Einflüsse der weiteren Eingangsgröße  $u_i$  und die Interaktionen mit allen zuvor einbezogenen Eingangsgrößen  $u_1, u_2, \dots, u_{i-1}$ . Alle Teilmodelle werden zu einem Gesamtmodell überlagert. Prinzipiell kann diese Überlagerung auf verschiedene Arten erfolgen, wie zum Beispiel durch eine Addition oder Multiplikation. In dieser Arbeit werden alle Teilmodelle additiv überlagert, da aus dieser Überlagerungsart eine interpretierbare Modellstruktur resultiert. Jedes Teilmodell bildet hierdurch die Haupteffekte der neuen Eingangsgröße  $u_i$  und die Interaktionen mit allen bisher betrachteten Ein-

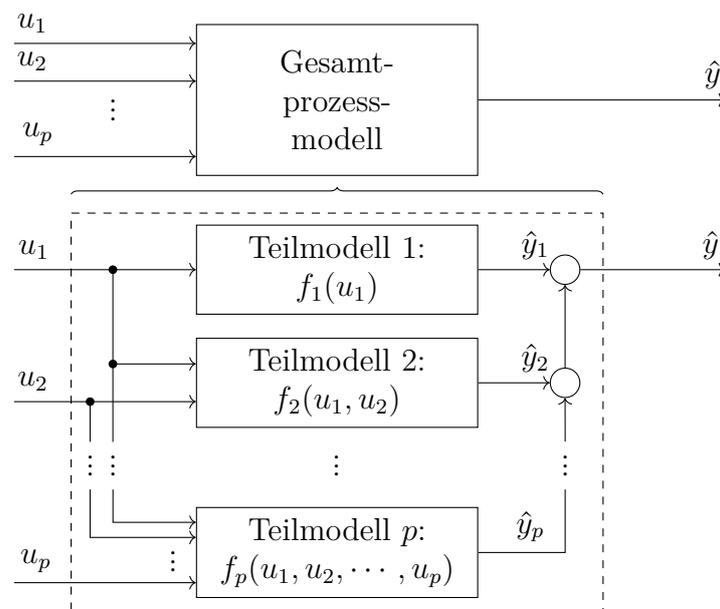


Bild 3.5: Struktur des Gesamtmodells bestehend aus additiv überlagerten Teilmodellen. In jedem der  $p$  Schritte wird ein neues Teilmodell hinzugefügt.

gangsgrößen  $u_1, u_2, \dots, u_{i-1}$  ab. Der Ausgang des Gesamtmodells

$$\hat{y} = \sum_{i=1}^p \hat{y}_i \quad (3.1)$$

ergibt sich aus der Addition der  $p$  Teilmodellausgänge  $\hat{y}_i$ . Mit diesem Konstruktionsprinzip liegt nach jedem Schritt ein nutzbares Prozessmodell vor, das einen Teil des Prozessverhaltens abbildet.

### Erstellung der Teilmodelle

In ILHAD wird jedes Teilmodell auf Basis eines individuellen Trainingsdatensatzes erstellt, der aus den Punkten des Versuchsplans  $\mathcal{D}_i$  im jeweiligen Schritt  $i$  und den gemessenen Ausgangswerten  $y(k)$  mit  $k = 1, 2, \dots, N_i$  besteht. Bild 3.6 stellt die Berechnung des Teilmodellfehlers  $e_i(k)$  dar, der als Basis für das Training des Teilmodells dient. Ein Versuchspunkt  $\underline{u}_i(k) = [u_1(k) \ u_2(k) \ \dots \ u_i(k)]^T \in \mathcal{D}_i$  gibt Werte für die  $i$  bisher betrachteten Eingangsgrößen vor. Alle weiteren  $M-i$  Eingangsgrößen werden mit  $\bar{\underline{u}}_{M-i} = [\bar{u}_{i+1} \ \bar{u}_{i+2} \ \dots \ \bar{u}_M]^T$  durch den Arbeitspunkt festgelegt. Somit ergibt sich der Eingangsgrößenvektor des Prozesses zu  $\underline{u}(k) = [\underline{u}_i(k) \ \bar{\underline{u}}_{M-i}]^T$ . Für die Modelle sind nur die Eingangsgrößen des Versuchspunkts  $\underline{u}_i(k)$  relevant.

Es wird angenommen, dass dem Prozessausgang ein additives weißes Rauschen  $n \sim \mathcal{N}(0, \sigma_n^2)$  mit der Rauschvarianz  $\sigma_n^2$  überlagert ist. Die Differenz zwischen dem gemessenen Prozessausgang  $y(k)$  und der Summe der Ausgänge aller bisherigen Teilmodelle  $\hat{y}_\Sigma(k) = \sum_{j=1}^{i-1} \hat{y}_j(k)$  ist die durch ein neues Teilmodell  $f_i$  zu modellieren-

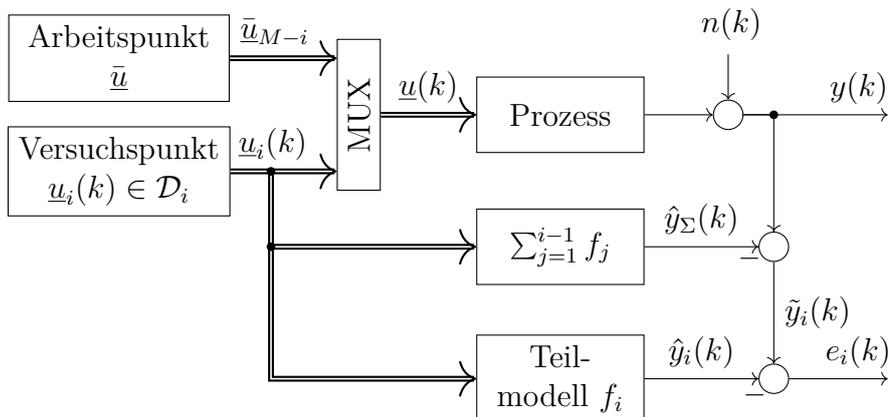


Bild 3.6: Berechnung des Teilmodellfehlers  $e_i$  [114]. Ein neues Teilmodell  $f_i$  bildet die Differenz  $\tilde{y}_i(k) = y(k) - \hat{y}_\Sigma(k)$  zwischen dem Prozessausgang  $y(k)$  und der Summe der Modellausgänge aller zuvor erstellten Teilmodelle  $\hat{y}_\Sigma(k)$  ab.

de Größe  $\tilde{y}_i(k) = y - \hat{y}_\Sigma(k)$ . Jedes Teilmodell wird hierzu auf Basis des Fehlers  $e_i(k) = \tilde{y}_i(k) - \hat{y}_i(k)$  trainiert, ohne die vorherigen Teilmodelle zu verändern. Somit kann ein Teilmodell auch als Fehlerkorrekturmodell interpretiert werden, das die aus der Variation der neuen Eingangsgröße resultierenden Fehler zwischen dem bisherigen Modell und dem Prozessverhalten bestmöglich korrigiert.

Damit weist ILHAD Parallelen zu *Boosting*-Verfahren auf, die ebenfalls Teilmodelle schrittweise additiv überlagern, um Fehler des zuvor erstellten Modells zu korrigieren. Ein bekanntes Verfahren ist das *Gradient Boosting* [33], dessen einfachste Variante die Summe der Fehlerquadrate als Verlustfunktion verwendet (L<sub>2</sub>-Boosting) und zumeist einfache Entscheidungsbäume als Teilmodelle verwendet. Mit diesem Prinzip können Boosting-Verfahren sehr hohe Modellgüten erzielen, wie es durch das populäre Verfahren *XGBoost* [18] demonstriert wird. Im Unterschied zum Boosting, bei dem alle Teilmodelle üblicherweise alle Eingangsgrößen verwenden, wird bei ILHAD in jedem Schritt eine weitere Eingangsgröße mit neuen Informationen für das Modell betrachtet. Dabei kann durch die gewählte Rangfolge der Eingangsgrößen angenommen werden, dass die Beiträge zusätzlicher Teilmodelle  $f_i$  zum Gesamtmodellausgang  $\hat{y}$  mit zunehmender Anzahl  $i$  an Schritten immer weiter abnehmen.

### 3.3 Modellstrukturen

Ein inkrementeller Versuchsplan ermöglicht durch seine in jedem Schritt raumfüllende Verteilung der Versuchspunkte die Verwendung unterschiedlicher Modellklassen für die Teilmodelle. Nachfolgend wird eine Anforderung an das Verhalten der Teilmodelle im zuvor betrachteten Unterraum definiert. Zusätzlich werden Szenarien zur Beurteilung und zum Vergleich der verschiedenen Modelle beschrieben.

#### Verhalten der Teilmodelle im zuvor betrachteten Unterraum

Bild 3.7 stellt exemplarisch einen inkrementellen Versuchsplan für den Schritt  $i = 2$  dar (Rauten), der zur Erstellung eines Modells  $f_2$  verwendet wird. Der Unterraum, in dem die Versuchspunkte aus dem vorherigen Schritt liegen, ist blau markiert. Das Prozessverhalten in diesem Unterraum wird bereits vollständig durch das Modell  $f_1$  abgebildet. In jedem ILHAD-Schritt ist das Prozessverhalten in einem solchen Unterraum  $\underline{u}_{i-1} = [u_1 \ u_2 \ \cdots \ u_{i-1}]^T$  mit  $u_j = \bar{u}_j \ \forall j = i, i+1, \dots, M$  bereits durch die vorherigen Modelle vollständig beschrieben. Im Gegensatz zu Boosting-Verfahren ist

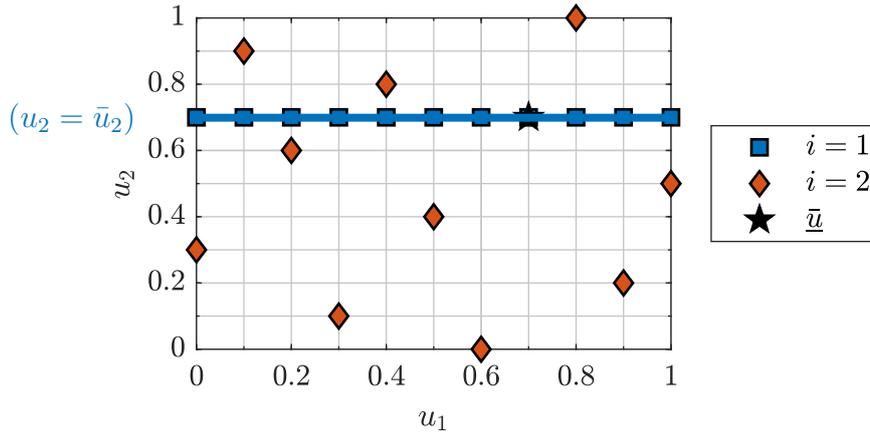


Bild 3.7: Inkrementeller Versuchsplan im Schritt  $i = 2$ . Der blau markierte Unterraum wird durch das Modell  $f_1$  modelliert. In diesem Bereich wird ein Modellausgang von exakt null von dem zu erstellenden Modell  $f_2$  gefordert.

bei ILHAD nicht zu erwarten, dass ein neues Modell Fehler der vorherigen Modelle korrigieren kann, da eine abnehmende Dichte der Versuchspunkte mit steigender Dimensionalität gegeben ist. Es wird daher gefordert, dass ein im  $i$ -ten Schritt additiv hinzugefügtes Teilmodell  $f_i$  keinen Einfluss auf den Gesamtmodellausgang im zuvor betrachteten Unterraum hat. Hierzu werden Teilmodellstrukturen verwendet, die einen Modellausgang von exakt null im Unterraum aufweisen. Auf diese Weise wird eine Verschlechterung der im Unterraum bereits erzielten Modellgüte in nachfolgenden Schritten verhindert.

### Szenarien zur simulativen Untersuchung

Um die ILHAD-Methodik mithilfe von Computersimulationen untersuchen zu können, werden geeignete Untersuchungsszenarien definiert. Die Beurteilung der Modellgüte erfolgt auf Basis verschiedener Testdatensätze, die unterschiedliche Bereiche des Eingangsraums abdecken. Da bei der Methodik ein vorgegebener Arbeitspunkt im Fokus steht, werden Testdatenverteilungen verwendet, die die Modellgüte in der Nähe dieses Punkts widerspiegeln. Hierzu wird mit

$$\mathcal{T} = \{(u_1, \dots, u_j, \dots, u_p, \bar{u}_{p+1}, \dots, \bar{u}_M) \mid 0 \leq u_j \leq 1 \forall j = 1, 2, \dots, p\},$$

mit  $u_j \sim \mathcal{N}(\bar{u}_j, \sigma_{\text{test}}^2)$ , (3.2)

eine um den Arbeitspunkt zentrierte Testdatenverteilung betrachtet, in der die ersten  $p$  Koordinaten aus je einer unabhängigen Normalverteilung mit der Standardabweichung  $\sigma_{\text{test}}$  erzeugt werden. Um Extrapolation zu vermeiden, wird der Wertebereich

auf das Intervall  $[0, 1]$  beschränkt. Punkte, bei denen mindestens eine Koordinate außerhalb dieses Intervalls liegt, werden verworfen und neu aus den Normalverteilungen generiert. Alle weiteren Koordinaten werden durch den Arbeitspunkt  $\bar{u}_{p+1}, \bar{u}_{p+2}, \dots, \bar{u}_M$  festgelegt. Bild 3.8 stellt verschiedene Testdatensätze mit jeweils  $N_{\text{test}} = |\mathcal{T}| = 100$  Datenpunkten in  $p = 2$  bzw.  $p = 3$  Dimensionen dar. Mit ansteigendem Wert von  $\sigma_{\text{test}}$  werden größere Bereiche des Eingangsraums gleichmäßiger abgedeckt. Für  $\sigma_{\text{test}} \rightarrow \infty$  ergibt sich eine Gleichverteilung, anstatt derer jedoch eine Sobol-Sequenz (vgl. Abschnitt 2.4.2) mit besseren raumfüllenden Eigenschaften zur Beurteilung der Modellgüte im gesamten Eingangsraum verwendet wird.

Um den Nutzen der inkrementellen Modellbildung bewerten zu können, werden Vergleichsmodelle benötigt. Für jede der betrachteten Modellklassen wird ein solches Vergleichsmodell in einem Schritt auf Basis eines gesonderten Versuchsplans erstellt. Hierzu wird ein mittels des EDLS-Algorithmus [26] optimierter LHC für alle  $p$  betrachteten Eingangsgrößen verwendet. Dieser LHC besitzt dieselbe Anzahl an Versuchspunkten  $N_{\Sigma} = \sum_{j=1}^p N_j$  wie ein inkrementeller Versuchsplan, bietet aber durch die Optimierung aller Punkte im  $p$ -dimensionalen Raum bessere raumfüllende Ei-

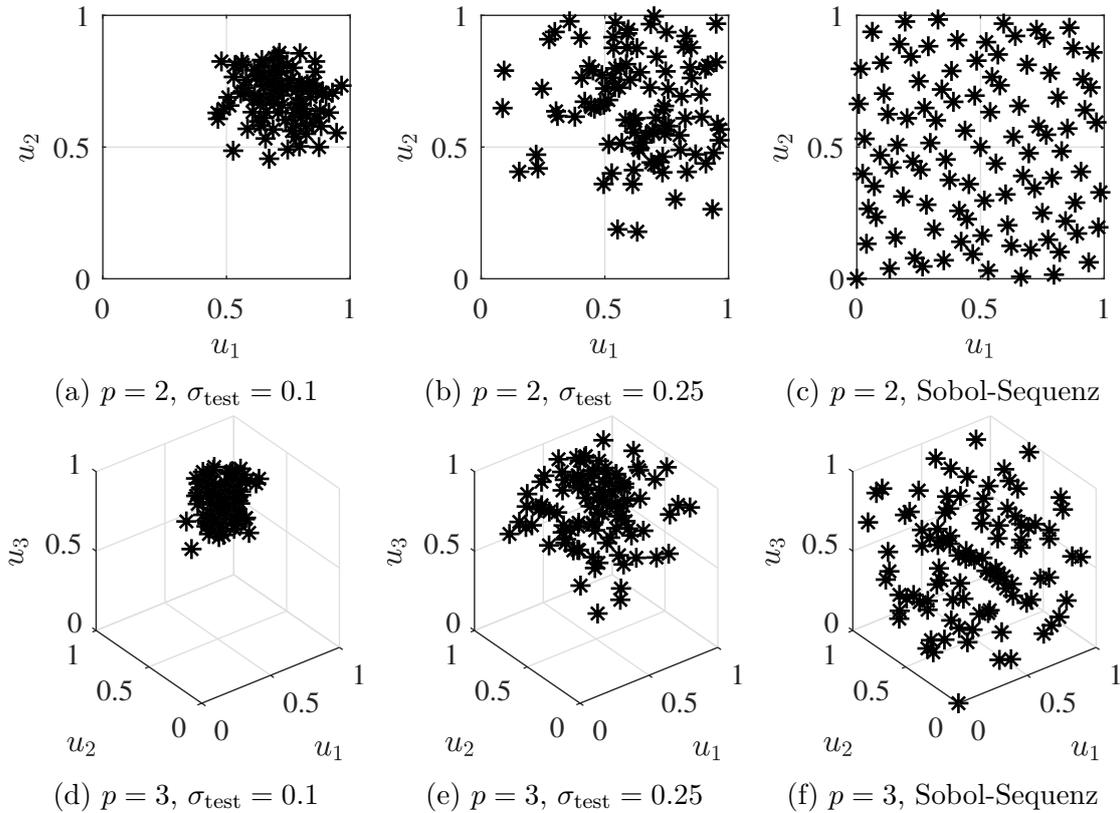


Bild 3.8: Darstellung verschiedener Testdatensätze mit jeweils  $N_{\text{test}} = 100$  Datenpunkten für  $p = 2$  mit  $\bar{u} = [0.7 \ 0.7]^T$  bzw.  $p = 3$  mit  $\bar{u} = [0.7 \ 0.7 \ 0.7]^T$ .

genschaften. Indem dieselbe Anzahl an Datenpunkten und möglichst vergleichbare Modellstrukturen mit z. B. demselben polynomialen Grad verwendet werden, werden die Auswirkungen des inkrementellen Versuchsplans und der inkrementellen Modellstruktur deutlich. Tabelle 3.1 stellt die Konstruktionsprinzipien von inkrementellem und Vergleichsmodell gegenüber.

Als zu modellierende Prozesse werden in diesem Kapitel verschiedene Testfunktionen betrachtet, die auf den Wertebereich  $[0, 1]$  normiert sind, um vergleichbare Rauscheinflüsse zu erzielen. Bild 3.9 stellt diese Testfunktionen mit  $M = 3$  betrachteten Eingangsgrößen dar, wobei  $u_3 = 0.7$  gewählt wurde und ausschließlich die ersten zwei Eingangsgrößen  $u_1$  und  $u_2$  variiert werden. Die multivariate Gaußfunktion in (a) ist in der Mitte des betrachteten Eingangsraums zentriert und weist durch eine zunehmende Standardabweichung mit steigendem Eingangsindex  $j = 1, 2, \dots, M$

Tabelle 3.1: Konstruktionsprinzipien von inkrementellem und Vergleichsmodell.

	Inkrementelles Modell	Vergleichsmodell
<b>Versuchsplanung</b>		
Art des Versuchsplans	inkr. Versuchsplan	optimierter LHC
Anzahl Versuchspunkte	$N = [N_1 \ N_2 \ \dots \ N_p]^T$	$N_\Sigma = \sum_{j=1}^p N_j$
Art der Optimierung	in jedem Schritt	in einem Schritt
<b>Modellbildung</b>		
Modellstruktur	additiv	einzelnes Modell
Modellklasse	Polynom / LMN / GPM	
Polynomialer Grad	$m = 2$	
LMN Fehlerschwelle	$\epsilon = 0.005$	

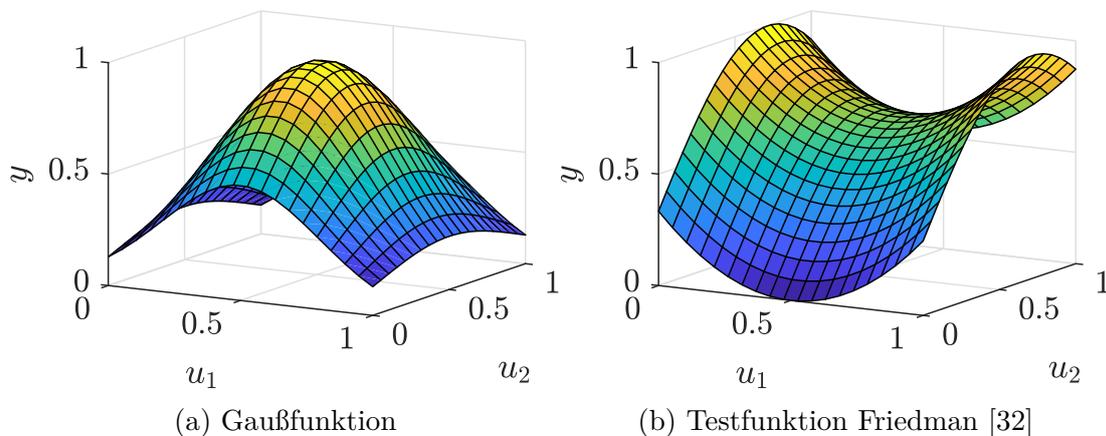


Bild 3.9: Darstellung der betrachteten Testfunktionen mit zwei variierten Eingangsgrößen und  $u_3 = 0.7$ .

eine eindeutige Rangfolge der Eingangsgrößen auf. Sie ist definiert durch

$$y(k) = \exp\left(-\frac{1}{2} \sum_{j=1}^M \frac{(u_j(k) - 0.5)^2}{(0.3 \cdot 1.5^j)^2}\right) \quad (3.3)$$

und beinhaltet Interaktionen zwischen allen Eingangsgrößen. Die zweite Testfunktion wurde in [32] beschrieben und ist für maximal  $M = 5$  Eingangsgrößen definiert. Zur Ermittlung einer Rangfolge werden ausgehend vom Arbeitspunkt die einzelnen Eingangsgrößen  $u_j \in [0, 1]$  mit  $j = 1, 2, \dots, 5$  variiert. Die Eingangsgröße, die die größte Spannweite der Ausgangsgröße  $y$  hervorruft, wird als erste in der Rangfolge betrachtet. Dies führt gegenüber der ursprünglichen Definition mit

$$y(k) = 20(u_1(k) - 0.5)^2 + 10 \sin(\pi u_2(k) u_3(k)) + 10u_4(k) + 5u_5(k) \quad (3.4)$$

zu den vertauschten Eingangsgrößen  $u_1$  und  $u_3$ . Die Funktion ist in Abhängigkeit der ersten zwei Eingangsgrößen in (b) dargestellt und besitzt sowohl Eingangsgrößen mit Interaktionen ( $u_2$  und  $u_3$ ) als auch Eingangsgrößen, die nur über ihre Haupteffekte wirken ( $u_1$ ,  $u_4$  und  $u_5$ ).

Die Simulationen unterliegen stochastischen Einflüssen durch das Prozessrauschen und die zufällige Initialisierung der verwendeten LHC. Um aussagekräftige Ergebnisse zu erhalten, sind daher alle Ergebnisse der Simulationen über 100 Wiederholungen gemittelt. Dabei werden folgende Standardparameter verwendet:

- Anzahl der ILHAD-Schritte  $p = M = 4$ ,
- Varianz des Prozessrauschens  $\sigma_n^2 = 1 \cdot 10^{-6}$ ,
- Versuchspunkte in den einzelnen Schritten  $\underline{N} = [31 \ 30 \ 30 \ 30]^T$ ,
- Arbeitspunkt  $\underline{u} = [0.7 \ 0.7 \ 0.7 \ 0.7]^T$ ,
- Anzahl der Testdatenpunkte  $N_{\text{test}} = 1000$ .

### 3.3.1 Polynommodelle

In diesem Abschnitt werden inkrementelle Modelle auf Basis von Polynommodellen (vgl. Abschnitt 2.2.1) beschrieben. Der Grad dieser Polynommodelle  $m \in \mathbb{N}$  kann dabei frei gewählt werden. Nachfolgend wird zur vereinfachten Darstellung ein Grad von  $m = 2$  verwendet. Das erste Teilmodell ergibt sich so zu

$$\hat{y}_1 = b_{0,1} + b_{1,1} \cdot u_1 + b_{2,1} \cdot u_1^2 \quad (3.5)$$

und bildet das Prozessverhalten in Abhängigkeit von  $u_1$  ab, solange alle weiteren Eingangsgrößen  $u_j = \bar{u}_j \forall j = 2, 3, \dots, M$  den Wert des Arbeitspunkts aufweisen.

### Erweiterung der Teilmodellstruktur

Die Polynommodelle in allen weiteren Schritten  $i \geq 2$  werden jeweils mit dem Linearfaktor  $(u_i - \bar{u}_i)$  multipliziert, der einen Modellausgang von null im zuvor betrachteten Unterraum erzwingt. Prinzipiell können zu diesem Zweck beliebige Funktionen genutzt werden, die für  $u_i = \bar{u}_i$  einen Wert von null aufweisen und daher im Weiteren als *Nullstellenfunktion* bezeichnet werden. In Verbindung mit Polynommodellen wird der Linearfaktor als Nullstellenfunktion verwendet, da dieser zu einer rein polynomialen Struktur der Teilmodelle führt. Das zweite Teilmodell

$$\begin{aligned} \hat{y}_2 = & \underbrace{(b_{0,2} + b_{1,2} \cdot u_1 + b_{2,2} \cdot u_2)}_{\text{Basismodell}} \cdot \underbrace{(u_2 - \bar{u}_2)}_{\text{Linearfaktor}} = b_{0,2} \cdot \underbrace{(u_2 - \bar{u}_2)}_{x_1} + b_{1,2} \cdot \underbrace{(u_1 u_2 - u_1 \bar{u}_2)}_{x_2} \\ & + b_{2,2} \cdot \underbrace{(u_2^2 - u_2 \bar{u}_2)}_{x_3} \end{aligned} \quad (3.6)$$

besteht somit aus einem polynomialen Basismodell, welches mit dem Linearfaktor multipliziert wird, und führt im dargestellten Fall zu drei Regressoren  $x_1, x_2, x_3$  mit dem Parametervektor  $\hat{\theta}_2 = [b_{0,2} \ b_{1,2} \ b_{2,2}]^T$ . Da die Multiplikation den maximalen Grad der Terme um eins erhöht, wird für das Basismodell der Grad  $m - 1$  gewählt. Auf diese Weise wird die Anzahl an zu schätzenden Parametern klein gehalten und höhere Wechselwirkungen der Eingangsgrößen im Modell vermieden, da diese in der Praxis oft vernachlässigt werden können [103]. Dieses Konstruktionsprinzip kann beliebig fortgesetzt werden, sodass sich beispielsweise im Schritt  $i = 3$  das Modell

$$\begin{aligned} \hat{y}_3 = & \underbrace{(b_{0,3} + b_{1,3} \cdot u_1 + b_{2,3} \cdot u_2 + b_{3,3} \cdot u_3)}_{\text{Basismodell}} \cdot \underbrace{(u_3 - \bar{u}_3)}_{\text{Linearfaktor}} = b_{0,3} \cdot (u_3 - \bar{u}_3) \\ & + b_{1,3} \cdot (u_1 u_3 - u_1 \bar{u}_3) + b_{2,3} \cdot (u_2 u_3 - u_2 \bar{u}_3) + b_{3,3} \cdot (u_3^2 - u_3 \bar{u}_3) \end{aligned} \quad (3.7)$$

mit  $i + 1 = 4$  Parametern  $\hat{\theta}_3 = [b_{0,3} \ b_{1,3} \ b_{2,3} \ b_{3,3}]^T$  ergibt. Das resultierende Gesamtmodell, welches sich gemäß Gleichung (3.1) aus der Addition aller Teilmodelle ergibt, entspricht so stets einem Polynommodell  $m$ -ten Grades.

Alle Teilmodelle sind linear in den Parametern, weshalb die Teilmodellparameter

$$\hat{\theta}_i = (\underline{X}_i^T \underline{X}_i)^{-1} \underline{X}_i^T \tilde{y}_i \quad (3.8)$$

in jedem  $i$ -ten Schritt mittels der Methode der kleinsten Fehlerquadrate (vgl. Gleichung (2.7)) bestimmt werden. Hierzu wird eine individuelle Regressormatrix  $\underline{X}_i$  und der Ausgangsgrößenvektor  $\tilde{\underline{y}}_i = [\tilde{y}_i(1) \ \tilde{y}_i(2) \ \cdots \ \tilde{y}_i(N_i)]^T$  verwendet. Die Regressormatrix ergibt sich aus den Regressoren des Teilmodells und den Versuchspunkten im jeweiligen inkrementellen Versuchsplan  $\mathcal{D}_i$ . Bild 3.10 stellt beispielhaft die auf diese Weise erstellten Teilmodelle in den Schritten (a)  $i = 1$  und (b)  $i = 2$  sowie in (c) die additive Überlagerung dieser Teilmodelle dar. Die Teilmodelle  $f_1$  und  $f_2$  bilden jeweils einen Teil des Prozessverhaltens ab und ergeben in Summe eine Approximation der Testfunktion Gauß für zwei Eingangsgrößen.

### Vergleich von raumfüllenden und D-optimalen Versuchsplänen

Die ILHAD-Methodik verwendet optimierte LHC als raumfüllende Versuchspläne (vgl. Abschnitt 3.2.1). Als Alternative werden in diesem Abschnitt D-optimale Versuchspläne sowohl in jedem Schritt als auch global für das Vergleichsmodell gegenübergestellt. D-optimale Versuchspläne (vgl. Abschnitt 2.4.1) basieren auf einer vorgegebenen Modellstruktur und platziert die Versuchspunkte so, dass der Konfidenzbereich des Parametervektors ein minimales Volumen aufweist [103]. Zu jedem inkrementellen Modell wird ein Vergleichsmodell erstellt, das mit einem Grad von  $m = 2$  dieselbe Struktur aufweist wie das aus der Addition der Teilmodelle resultierende Gesamtmodell. Somit sind die inkrementellen Modelle und die Vergleichsmodelle bezüglich der Anzahl an Parametern und dem polynomialen Grad vergleichbar.

In Bild 3.11 werden mehrere inkrementelle Polynommodelle und polynomiale Vergleichsmodelle mithilfe verschiedener Testdatensätze beurteilt. Diese Testdatensätze

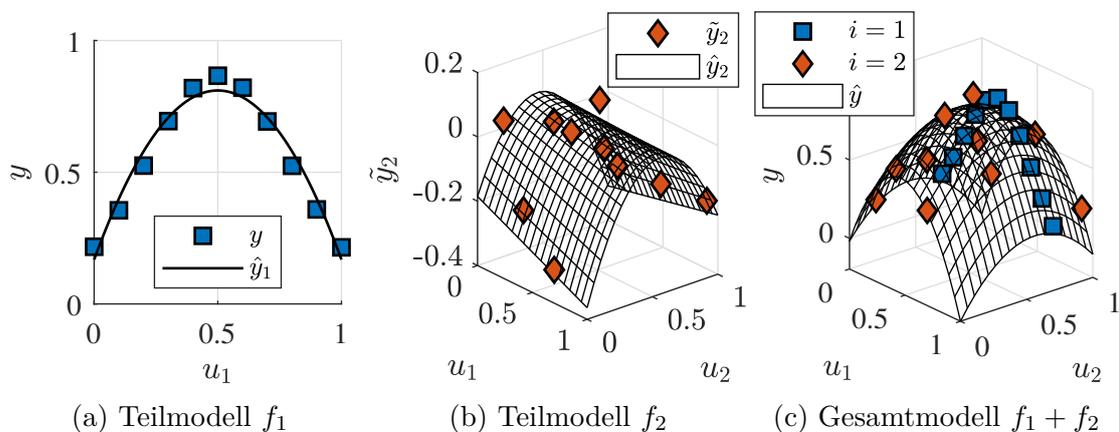
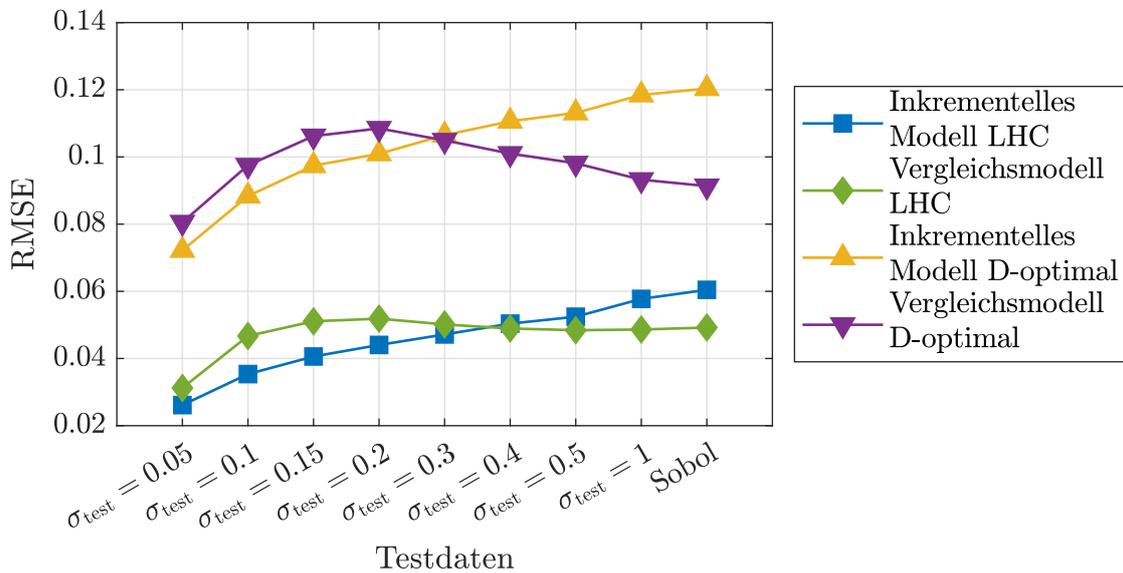


Bild 3.10: Polynomiale Teilmodelle und resultierendes Gesamtmodell für  $p = 2$  und  $\underline{N} = [11 \ 10]^T$  am Beispiel der Gaußfunktion gemäß Gleichung (3.3).



(a) Gaußfunktion gemäß Gl. (3.3) als Testfunktion

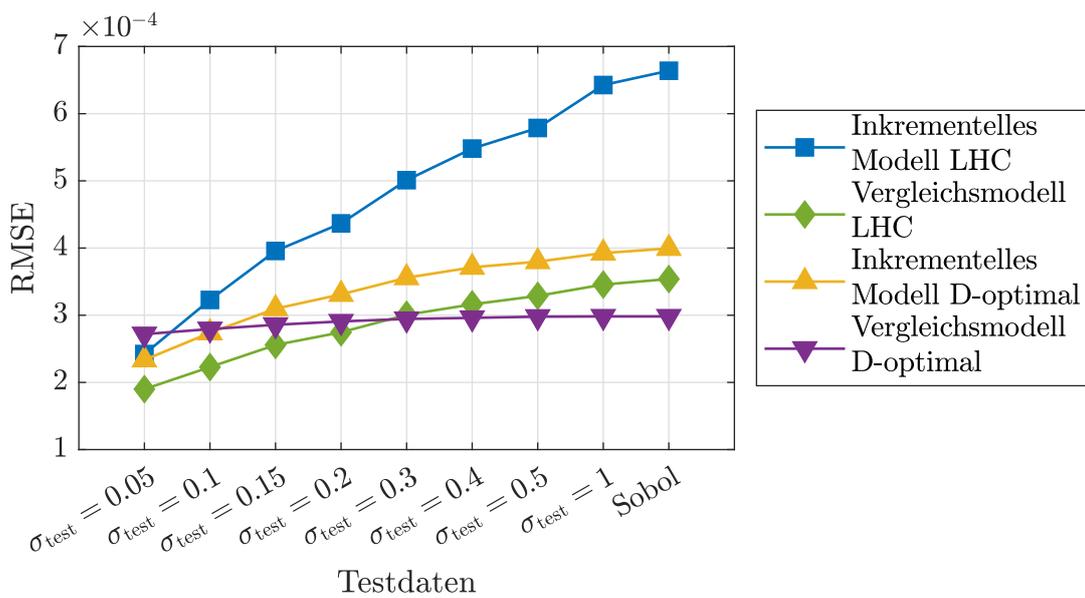
(b) Polynom  $f(\underline{u}) = 7u_1^2 + 5u_2^2 + 3u_3 + u_4 + 4u_1u_3 + 3u_2u_4 + 6$  als Testfunktion

Bild 3.11: Vergleich des RMSE für verschiedene Polynommodelle. Die inkrementellen Modelle und die Vergleichsmodelle besitzen jeweils  $p = 4$  Eingangsgrößen und werden mit verschiedenen Testdatensätzen evaluiert, die von links nach rechts den Eingangsraum gleichförmiger abdecken.

decken ausgehend vom Arbeitspunkt unterschiedliche Bereiche des Eingangsraums ab, wobei von links nach rechts eine zunehmende Raumabdeckung gegeben ist.

In Bild 3.11a wird die Gaußfunktion gemäß Gleichung (3.3) als Testfunktion betrachtet. Beide inkrementelle Modelle erzielen in der Nähe des Arbeitspunkts ( $\sigma_{\text{test}} < 0.3$ ) eine höhere Güte als das zugehörige Vergleichsmodell. Im gesamten Eingangsraum weisen jedoch die Vergleichsmodelle durch die freie Platzierung der Versuchspunkte einen geringeren Fehler auf. Die mittels optimierter LHC erzeugten raumfüllenden Versuchspläne führen bei dieser Testfunktion im Vergleich zu den D-optimalen Versuchsplänen zu geringeren Werten des RMSE. So sind die D-optimalen Versuchspläne zwar in Bezug auf die angenommene polynomiale Modellstruktur optimal, jedoch entspricht diese nicht dem funktionalen Zusammenhang der betrachteten Testfunktion. Hiermit liegt ein dominanter Biasfehler vor, der durch die Verwendung raumfüllender Versuchspläne reduziert werden kann.

Stimmt hingegen die angenommene Modellstruktur der D-optimalen Versuchspläne mit der Testfunktion überein, werden die Vorteile dieser Art von Versuchsplänen deutlich. Bild 3.11b stellt diesen Fall mit einem quadratischen Polynom als Testfunktion dar. Das inkrementelle Modell auf Basis der D-optimalen Versuchspläne weist für alle Testdatensätze geringere Fehler auf als das LHC basierte inkrementelle Modelle. Auch das Vergleichsmodell erzielt für  $\sigma_{\text{test}} > 0.2$  bessere Ergebnisse mit dem D-optimalen Versuchsplan.

Aufgrund der zumeist unbekannt funktionalen Zusammenhänge von realen Prozessen ist die Wahl einer abweichenden polynomialen Modellstruktur eine für die Praxis realistische Annahme, die zu einem dominierenden Biasfehler führt. Wie exemplarisch aufgezeigt können raumfüllende Versuchspläne in solchen Fällen zu geringeren Gesamtfehlern führen als modellbasierte Versuchspläne. Dies bestätigt die Wahl von raumfüllenden Versuchsplänen im Rahmen der ILHAD-Methodik.

### 3.3.2 Lokale Modellnetze

Wird die inkrementelle Modellbildung auf Prozesse mit ausgeprägtem nichtlinearem Verhalten angewandt, sind die zuvor betrachteten polynomialen Modelle oft unzureichend. Im Folgenden wird die flexiblere Modellklasse der lokalen Modellnetze (LMN) verwendet. Die Grundlagen dieser Modellklasse sowie der zum Training eingesetzte LOLIMOT-Algorithmus sind in Abschnitt 2.2.2 beschrieben.

In der inkrementellen Modellstruktur gemäß Bild 3.5 werden die einzelnen Teilmodelle  $f_i$  nacheinander in jedem Schritt  $i = 1, 2, \dots, p$  mit dem LOLIMOT-Algorithmus erstellt. Jedes Teilmodell besteht aus  $L_i$  lokalen Modellen  $g_{i,l}(\underline{u})$  mit  $l = 1, 2, \dots, L_i$ . Standardmäßig verwendet LOLIMOT lokal affine Modelle, die auch im Schritt  $i = 1$  der inkrementellen Modellbildung genutzt werden. Hiermit ergeben sich die  $l = 1, 2, \dots, L_1$  lokalen Modelle im ersten Schritt zu

$$g_{1,l}(u_1) = b_{0,l} + b_{1,l} \cdot u_1. \quad (3.9)$$

### Erweiterung der Struktur der lokalen Modelle

Um die in Abschnitt 3.3 beschriebene Anforderung an das Teilmodellverhalten im zuvor betrachteten Unterraum zu erfüllen, wird jedes lokale Modell in den Schritten  $i \geq 2$  analog zu den Polynomen in Abschnitt 3.3.1 erweitert. Hierzu ergibt sich mit

$$g_{i,l}(u_1, u_2, \dots, u_i) = \underbrace{\left( b_{0,l} + \sum_{j=1}^i b_{j,l} \cdot u_j \right)}_{\text{affines Basismodell}} \cdot \underbrace{\Psi(u_i - \bar{u}_i)}_{\text{Nullstellenfunktion}} \quad (3.10)$$

jedes der  $l = 1, 2, \dots, L_i$  lokalen Modelle aus der Multiplikation eines affinen Basismodells mit einer Nullstellenfunktion  $\Psi(u)$ . Die Nullstellenfunktion ist eine stetige Funktion für die  $\Psi(0) = 0$  gilt, sodass durch das Argument  $u_i - \bar{u}_i$  ein Modellausgang von null im zuvor betrachteten Unterraum erzwungen wird. Da die lokalen Modelle trotz dieser Erweiterung linear in den Parametern sind, lassen sich die Modellparameter mit der gewichteten Methode der kleinsten Quadrate gemäß Gleichung (2.13) schätzen.

Nachfolgend werden zwei alternative Nullstellenfunktionen betrachtet. Analog zu den Polynommodellen in Abschnitt 3.3.1 wird mit  $\Psi(u) = u$  ein Linearfaktor verwendet. Die zweite Nullstellenfunktion basiert auf einem quadratischen B-Spline. Ein B-Spline ist über Knoten definiert, zwischen denen stets positive Funktionswerte mittels abschnittsweise gültiger Polynome erzeugt werden. An den Knoten, an denen diese Polynome aufeinandertreffen, liegen zusätzliche Bedingungen bezüglich Stetigkeit und Differenzierbarkeit vor. Ein quadratischer B-Spline verwendet abschnittsweise quadratische Polynome, die zu einem stetigen Funktionsverlauf mit stetiger erster Ableitung führen [22]. Der hier verwendete B-Spline  $B(u) = B(u|\underline{k})$  ist über den Knotenvektor  $\underline{k} = [(-1/2 \cdot \delta) \ (-1/4 \cdot \delta) \ (1/4 \cdot \delta) \ (1/2 \cdot \delta)]$  definiert. Der Parameter  $\delta$  gibt die Breite der Funktion vor, da definitionsgemäß außerhalb der Knoten

$B(u) = 0, u \notin [(-1/2 \cdot \delta), (1/2 \cdot \delta)]$  gilt. Hiermit ergibt sich die Nullstellenfunktion

$$\Psi(u) = 1 - \frac{B(u)}{B(0)}. \quad (3.11)$$

Beide Nullstellenfunktionen sind in Bild 3.12 grafisch gegenübergestellt. Der Linearfaktor weist eine konstante Steigung über den Eingangsraum auf und beeinflusst so alle lokalen Modelle. Der Einfluss des B-Splines ist hingegen lokal begrenzt und betrifft hauptsächlich lokale Modelle mit Partitionen im Bereich des Arbeitspunkts. Lokale Modelle mit großem Abstand zum Arbeitspunkt werden somit kaum verändert.

Bild 3.13 stellt die Modellausgänge des Modells  $f_2$  mit verschiedenen Nullstellenfunktionen gegenüber, wobei der im Weiteren verwendete Wert  $\delta = 0.2$  für die Breite des B-Splines dargestellt ist. Während in (a) durch den Linearfaktor ein quadratisches Modellverhalten zu sehen ist, bleiben in (b) die lokal affinen Modelle bestehen und werden ausschließlich im Bereich um die Koordinate des Arbeitspunkts  $u_2 = \bar{u}_2 = 0.7$  verändert. Auf diese Weise bleibt in (b) für weit vom Arbeitspunkt entfernte Modelle die lokale Interpretierbarkeit und das Extrapolationsverhalten erhalten.

### Transformation des Versuchsplans

Ein Vorteil im inkrementellen Vorgehen besteht in der Möglichkeit zur Nutzung von modelliertem Prozessverhalten aus vorherigen Schritten. Bisher erlernte Zusammenhänge zwischen  $y$  und  $u_1, u_2, \dots, u_{i-1}$  liegen in Form von Modellparametern und Partitionen der LMN vor. Die Partitionen charakterisieren hierbei die Nichtlinearitäten des Prozesses. Wird ein LMN mittels des LOLIMOT-Algorithmus erstellt, führen

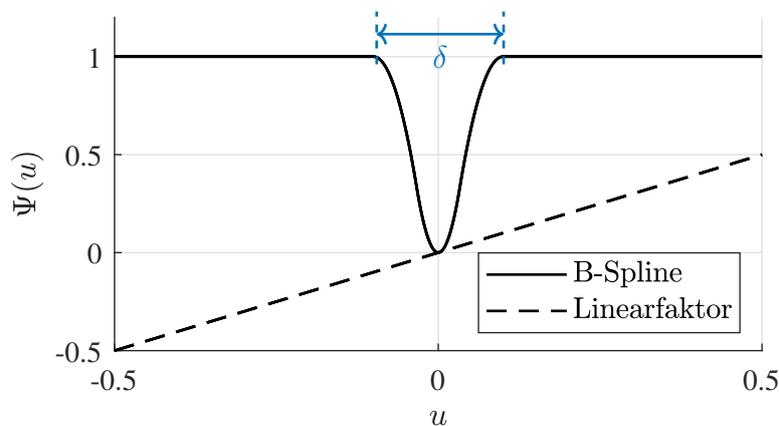


Bild 3.12: Verläufe der verwendeten Nullstellenfunktionen [113].

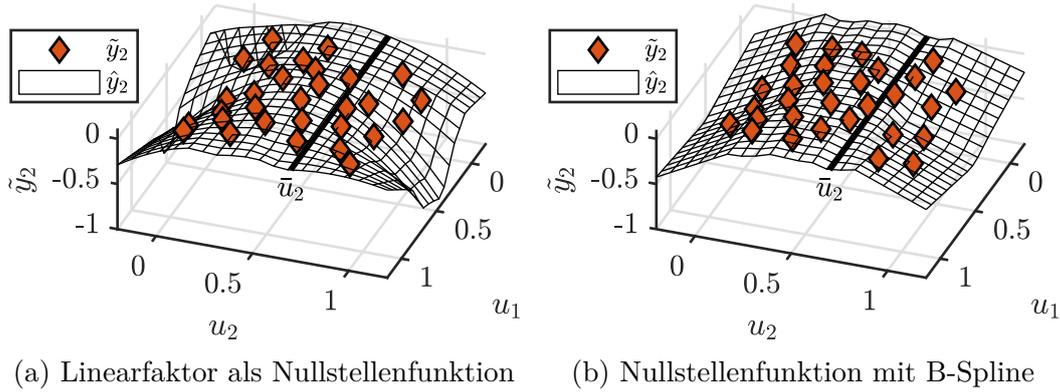


Bild 3.13: Vergleich des Teilmodells  $f_2$  am Beispiel der Testfunktion Gauß gemäß Gleichung (3.3) mit verschiedenen Nullstellenfunktionen  $\Psi(u_2 - \bar{u}_2)$ .

Nichtlinearitäten in Bereichen des Eingangsraums zu einer feineren Partitionierung dieser Bereiche [80]. Nachfolgend wird ein Verfahren beschrieben, dass die Partitionierung der bestehenden LMN zur Verbesserung des inkrementellen Versuchsplans nutzt. Ziel ist es, die Dichte an Versuchspunkten in stark nichtlinearen Bereichen des Eingangsraums zu steigern, wozu der inkrementelle Versuchsplan transformiert wird. Ein neues Teilmodell kann so diese Bereiche mit komplexem Prozessverhalten besser abbilden.

In jedem Schritt  $i \geq 2$  wird der inkrementelle Versuchsplan gemäß Abschnitt 3.2.1 um neue Versuchspunkte  $\mathcal{D}_i$  erweitert. Diese Versuchspunkte bilden zunächst einen LHC mit äquidistanten Stufen. Mittels der Transformation  $\tilde{u}_j = z_j(u_j)$ , die für jede bereits betrachtete Eingangsgröße  $u_j$ , mit  $j = 1, 2, \dots, i - 1$  durchgeführt wird, werden diese Stufen an die zuvor erkannten Nichtlinearitäten angepasst. Die Transformationsfunktion  $z_j(u_j)$  ist eine abschnittsweise lineare Funktion und ergibt sich aus der Partitionierung des Teilmodells  $f_j$ , das  $u_j$  als neue Eingangsgröße in das Gesamtmodell einbezieht und deren Haupteffekte abbildet.

Bild 3.14 zeigt exemplarisch, wie sich diese Transformationsfunktion (durchgezogene Linie) aus der Partitionierung des Teilmodells  $f_j$  ergibt. Hierzu werden die Grenzen aller  $L_j$  Partitionen in dem jeweiligen Modell  $f_j$  auf die Achse  $u_j$  projiziert (linke Seite). Jede der  $\tilde{L}_j \leq L_j$  projizierten Partitionen besitzt eine Ausdehnung  $d_k$  mit  $k = 1, \dots, \tilde{L}_j$  und definiert einen Abschnitt der Transformationsfunktion (blaue Markierungen), in dem der Funktionswert linear um den Wert  $d_k$  ansteigt. Die Breite jedes Abschnitts wird mit  $\Delta d_k$  bezeichnet und mit der Normierung

$$\Delta \tilde{d}_k = \frac{\Delta d_k}{\sum_{l=1}^{\tilde{L}_j} \Delta d_l} \quad (3.12)$$

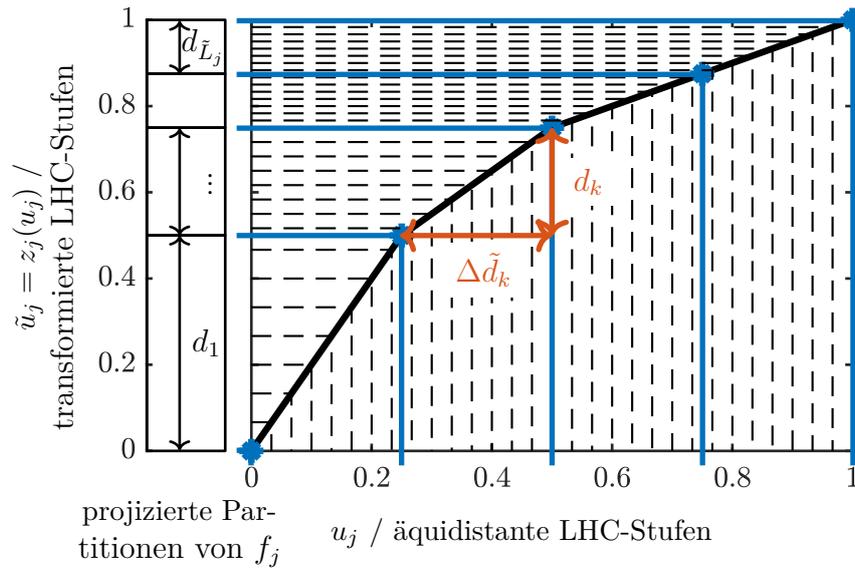


Bild 3.14: Konstruktion der abschnittsweise linearen Transformationsfunktion  $z_j(u_j)$  aus den projizierten Partitionen des Modells  $f_j$ , wobei jeder Abschnitt  $k$  über die Ausdehnung  $d_k$  und Breite  $\Delta \tilde{d}_k = 1/4$  definiert ist. Die transformierten Stufen (links) ergeben sich aus den äquidistanten Stufen (unten).

an den vorgegebenen Wertebereich der Eingangsgrößen  $u_j \in [0, 1] \forall j = 1, 2, \dots, M$  angepasst. Analytisch lassen sich die einzelnen Abschnitte  $k$  durch

$$z_j(u_j) = \frac{d_k}{\Delta \tilde{d}_k} \left( u_j - \sum_{l=1}^k \Delta \tilde{d}_l \right) + \sum_{l=1}^k d_l, \quad \text{für} \quad \sum_{l=1}^{k-1} \Delta \tilde{d}_l \leq u_j < \sum_{l=1}^k \Delta \tilde{d}_l \quad (3.13)$$

beschreiben. In Bild 3.14 wird mit  $\Delta d_k = 1$  eine konstante Breite der Abschnitte betrachtet, sodass sich  $\Delta \tilde{d}_k = 1/\tilde{L}_j = 1/4$  ergibt. Beispielhafte LHC-Stufen sind als gestrichelte Linien dargestellt. Die äquidistanten Stufen (unten) werden zu Stufen, deren Dichte antiproportional zur Ausdehnung der Partitionen ist (links).

Zusätzlich kann die Breite der Abschnitte mit

$$\Delta d_k = d_k^r \quad (3.14)$$

in Abhängigkeit der Ausdehnung der projizierten Partitionen  $d_k$  gewählt werden, wobei der Exponent  $r$  die Transformationsstärke beeinflusst. Bild 3.15 stellt Transformationsfunktionen für verschiedene Werte von  $r$  bei einer gegebenen Partitionierung dar. Für  $r = 1$  ergibt sich eine Gerade, sodass die äquidistanten LHC-Stufen erhalten bleiben. Wird  $r = 0$  gewählt, ergibt sich die zuvor betrachtete konstante Breite  $\Delta d_k = 1$ . Für kleinere Werte wie beispielsweise  $r = -1$  wird die Transformation ver-

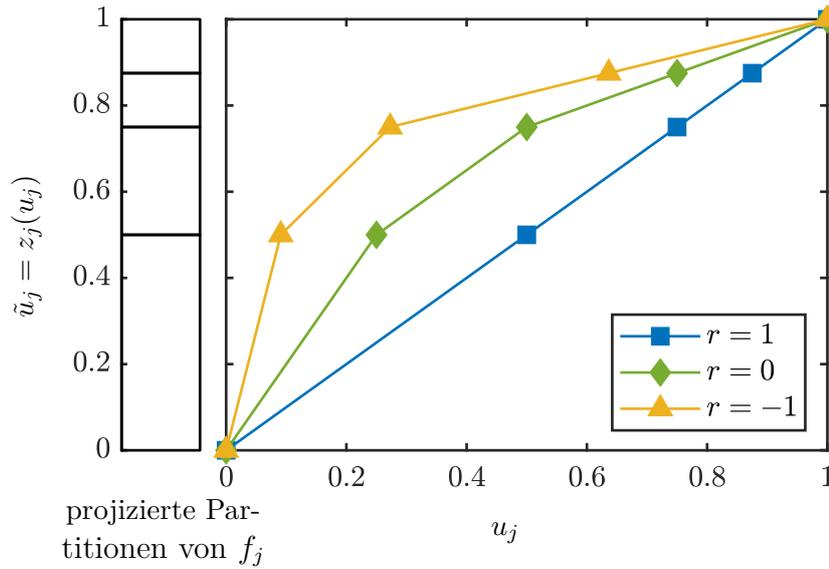


Bild 3.15: Transformationsfunktionen für eine gegebene Partitionierung (links) und unterschiedliche Exponenten  $r$ .

stärkt und somit die Dichte an LHC-Stufen in Bereichen mit kleineren Partitionen weiter gesteigert. Die Wahl des Exponenten  $r$  beeinflusst, in welchem Umfang Prozesswissen, das in Form der Partitionen gespeichert ist, in der Versuchsplanung berücksichtigt wird. Eine schwache Transformation  $r \in [0, 1)$  verwendet dieses Wissen kaum und hat primär zum Ziel, den Eingangsraum zu erkunden und gleichmäßig mit Versuchspunkten abzudecken. Eine starke Transformation  $r < 0$  nutzt das Vorwissen aus, um nichtlineare Bereiche besser abbilden zu können. Dies kann jedoch durch die verschlechterte Raumabdeckung zu größeren Fehlern in anderen Bereichen des Eingangsraums führen. Ein Parameter mit einem ähnlichen Effekt existiert im aktiven Lernen und beeinflusst dort den Grad der Erkundung des Prozessverhaltens (*Exploration*) und der Ausnutzung der erlernten Zusammenhänge (*Exploitation*) [109]. Im aktiven Lernen wird dieser Parameter so variiert, dass mit fortschreitender Lerndauer mehr Prozesswissen genutzt wird, während für die Transformation der Versuchspläne in der inkrementellen Modellbildung eine Festlegung im Vorfeld erfolgen muss.

### Demonstration an einem Testprozess

Um die Eigenschaften der in diesem Abschnitt beschriebenen inkrementellen Modellbildung auf Basis von LMN zu untersuchen, wird diese Methode auf die zwei in Abschnitt 3.3 beschriebenen Testfunktionen angewandt. Dabei werden mit  $r \in \{-1, -0.5, 0, 0.5, 1\}$  verschieden starke Transformationen der LHC-Stufen verglichen

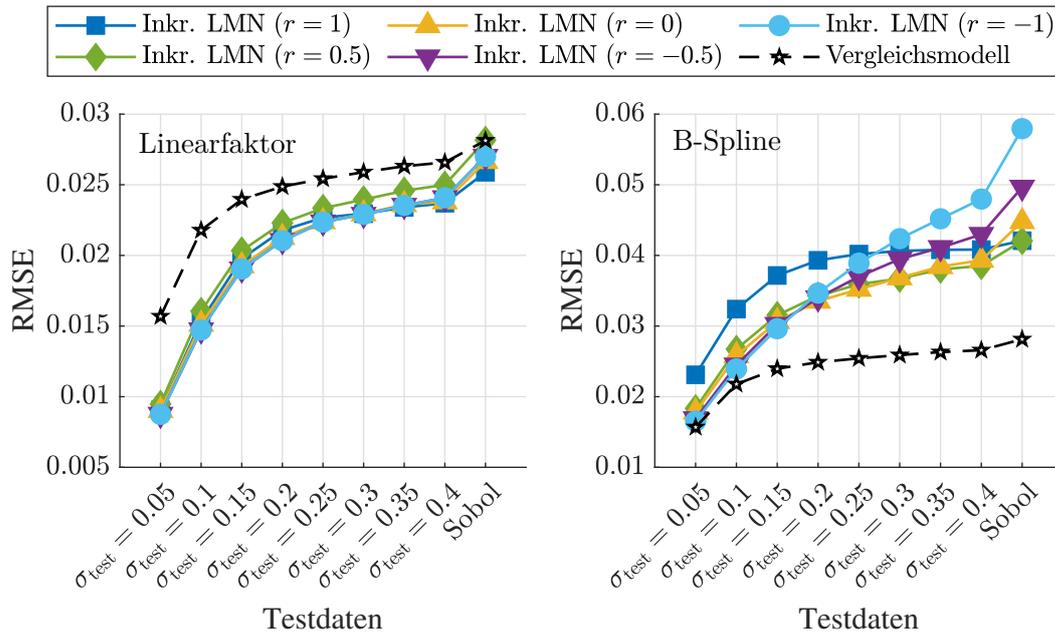
und zugleich mit dem Linearfaktor und der auf einem B-Spline basierenden Funktion verschiedene Nullstellenfunktionen betrachtet. Für alle Teilmodelle und die Vergleichsmodelle wird die Fehlerschwelle  $\epsilon = 0.005$  als Abbruchkriterium verwendet.

Bild 3.16 stellt den RMSE für die verschiedenen Modelle auf Basis unterschiedlich verteilter Testdaten gegenüber. Folgende drei wesentliche Beobachtungen lassen sich in Bild 3.16a anhand der Testfunktion Friedman machen:

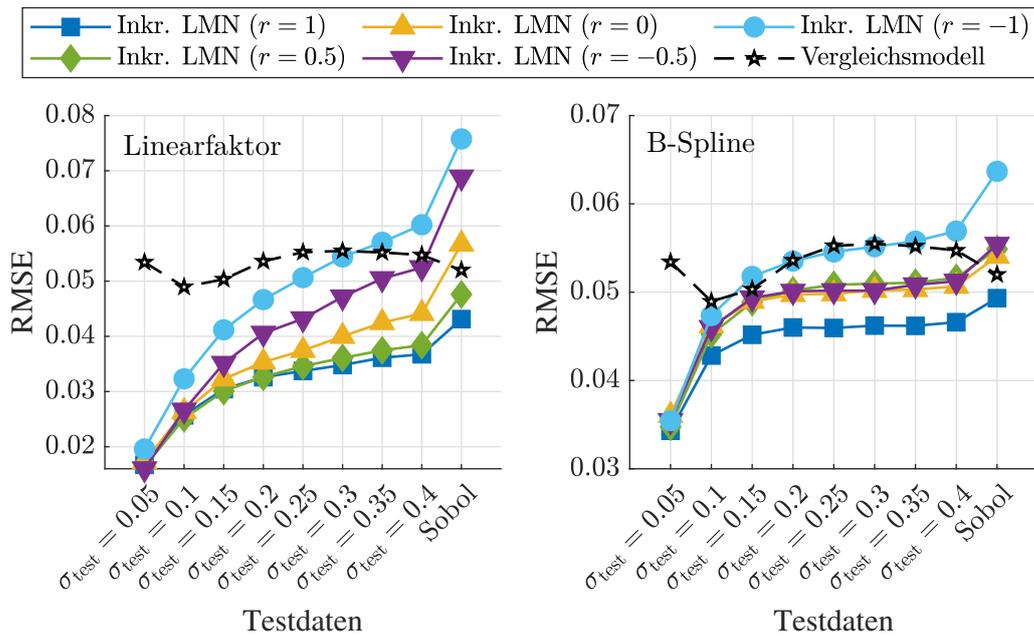
1. Durch eine Transformation kann die Modellgüte in der Nähe des Arbeitspunkts mit  $\sigma_{\text{test}} \leq 0.3$  gegenüber dem nicht transformierten Fall ( $r = 1$ ) gesteigert werden, wobei die besten Ergebnisse in diesem Bereich mit der stärksten Transformation  $r = -1$  erzielt werden. Die Modelle mit Transformation können von der gesteigerten Punktdichte in nichtlinearen Bereichen des Eingangsraums profitieren und dort eine größere Anzahl an lokalen Modellen erstellen.
2. Eine starke Transformation kann gegenüber dem nicht transformierten Versuchsplan ( $r = 1$ ) zu einem Anstieg des Fehlers im gesamten Eingangsraum (Sobol-Sequenz) führen. Dies kann durch die ungleichmäßige Datenverteilung erklärt werden, die die Abbildung von Interaktionen erschwert.
3. Insgesamt werden mit dem Linearfaktor (links) geringere Fehler erzielt als mit der auf einem B-Spline basierten Nullstellenfunktion (rechts), was auf die quadratischen Terme zurückzuführen ist, die aus der Multiplikation mit dem Linearfaktor resultieren. Die betrachtete Testfunktion kann durch diese Terme mit weniger lokalen Modellen genauer approximiert werden.

Für Prozesse mit ausgeprägten Interaktionen zwischen den Eingangsgrößen wie der Testfunktion Gauß kann hingegen eine Transformation nachteilig sein. So steigt der Fehler der inkrementellen Modelle in Bild 3.16b mit zunehmender Transformationsstärke an. Für beide Nullstellenfunktionen werden bei dieser Testfunktion die besten Ergebnisse mit den inkrementellen Modellen ohne Transformation ( $r = 1$ ) erzielt.

Für beide Test- und Nullstellenfunktionen können inkrementelle Modelle eine ähnliche oder höhere Modellgüte in der Nähe des Arbeitspunkts ( $\sigma_{\text{test}} \leq 0.05$ ) erzielen wie die Vergleichsmodelle. Außer mit der B-Spline basierten Nullstellenfunktion in Bild 3.16a existieren jeweils inkrementelle Modelle, die für alle Testdatensätze bessere Ergebnisse erzielen als das Vergleichsmodell. Es zeigt sich allerdings auch, dass der Nutzen einer Transformation von den Eigenschaften des betrachteten Prozesses abhängt. So kann die Modellgüte in der Nähe des Arbeitspunkts insbesondere dann durch die Transformation gesteigert werden, wenn ein Prozess nichtlineares Verhalten mit geringen Interaktionen zwischen den Eingangsgrößen aufweist.



(a) Testfunktion Friedman gemäß Gleichung (3.4)



(b) Testfunktion Gauß gemäß Gleichung (3.3)

Bild 3.16: Vergleich verschiedener Modelle auf Basis von LMN mit dem Linearfaktor (links) und der B-Spline basierten Nullstellenfunktion (rechts) bei verschiedenen Transformationsstärken  $r$ . Der RMSE wird anhand von Testdatensätzen bestimmt, die von links nach rechts den Eingangsraum zunehmend gleichmäßiger abdecken.

### 3.3.3 Gaußprozessmodelle

Mit Gaußprozessmodellen (GPM) wird eine weitere Modellklasse im Rahmen der inkrementellen Modellbildung betrachtet. GPM besitzen als nichtparametrische Modelle eine hohe Flexibilität und sind gut für kleine Trainingsdatenmengen geeignet, wie sie in Versuchen häufig gegeben sind. Die Grundlagen dieser Modellklasse sind in Abschnitt 2.2.3 zusammengefasst. Eine Erweiterung von GPM hin zu einer additiven Modellstruktur wurde bereits in [25] beschrieben, wobei verschiedene Untermengen der Eingangsgrößen in die Kernel-Funktionen einbezogen werden. Im Gegensatz zu ILHAD erfolgt die Modellerstellung allerdings in einem Schritt. Eine Strategie, um ein einzelnes GPM inkrementell zu erweitern wird in [110] vorgeschlagen, indem die Kovarianzmatrix nacheinander um die einzelnen Eingangsgrößen erweitert wird.

Nachfolgend wird analog zur Verwendung von Polynommodellen und LMN in der inkrementellen Modellstruktur in jedem Schritt ein separates GPM erstellt. Die Hyperparameter dieser GPM werden in jedem Schritt individuell durch eine Maximierung der Log-Marginal-Likelihood bestimmt (siehe Abschnitt 2.2.3). Eine Ausnahme ist die angenommene Standardabweichung des Rauschens  $\hat{\sigma}_n$ , die nur einmalig im ersten Schritt bestimmt wird. In diesem ILHAD-Schritt ist die höchste Dichte an Versuchspunkten gegeben, sodass von einer bestmöglichen Schätzung dieses Hyperparameters ausgegangen werden kann. Ein additives weißes Rauschen am Prozessausgang ist unabhängig von der Anzahl an variierten Eingangsgrößen und damit unabhängig vom ILHAD-Schritt. Der durch das Teilmodell  $f_1$  bestimmte Wert für  $\hat{\sigma}_n$  wird somit für alle weiteren Teilmodelle  $f_j$  mit  $j = 2, 3, \dots, p$  übernommen. Um zu verhindern, dass ein Teilmodell auf Basis von GPM das Gesamtmodell im zuvor betrachteten Unterraum verändert (vgl. Abschnitt 3.3), werden zwei verschiedene Methoden vorgestellt, die zusätzliche Dummy-Punkte in diesem Unterraum platzieren (GPM-DP) oder nicht stationäre Kernel-Funktionen verwenden (GPM-NS). Diese Ansätze treten an Stelle der Nullstellenfunktionen.

#### GPM mit Dummy-Punkten

Eine Methode, um einem GPM ein gewünschtes Verhalten aufzuprägen, ist in [12] beschrieben. Diese Methode verwendet ein Gitter aus Dummy-Datenpunkten (DP), um ein vorgegebenes Monotonieverhalten zu erzwingen. Die zu diesen Punkten gehörenden Ausgangswerte werden nicht vermessen, sondern durch ein Optimierungsverfahren bestimmt, wobei das geforderte Monotonieverhalten als Nebenbedingungen

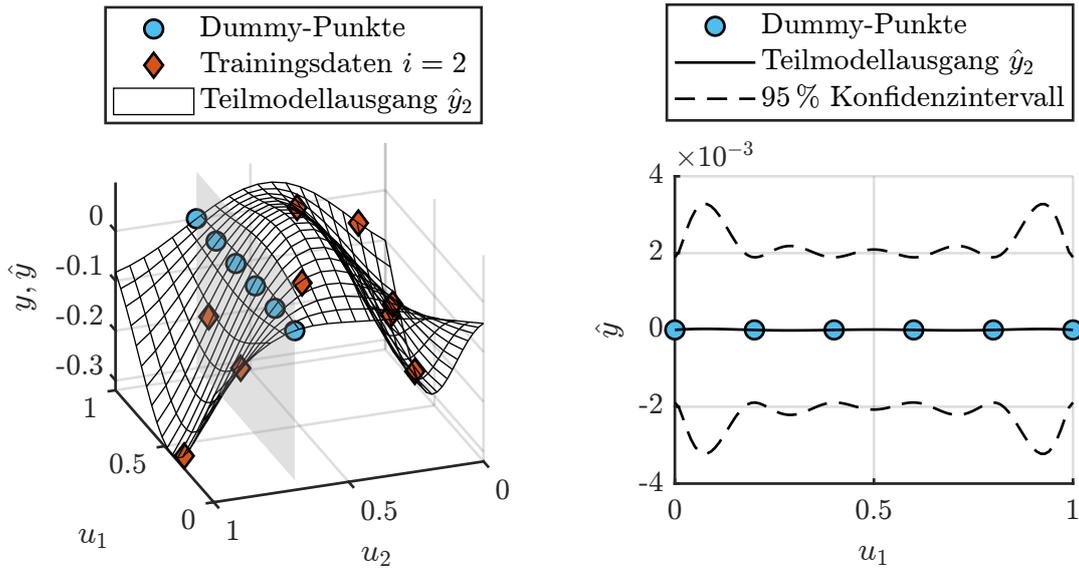
einfließt. Die in diesem Abschnitt beschriebene Teilmodellstruktur GPM-DP basiert auf derselben Grundidee und platziert in jedem  $i$ -ten Schritt  $N_{d,i}$  zusätzliche Dummy-Punkte  $\underline{u}_{d,i}(j)$  mit  $j = 1, 2, \dots, N_{d,i}$  im zuvor betrachteten Unterraum. Da für diesen Unterraum ein Teilmodellausgang von null gefordert ist, wird allen Dummy-Punkten der Ausgangswert null zugewiesen, sodass im Gegensatz zu [12] keine rechenaufwändige Optimierung der Ausgangswerte notwendig ist. Die Platzierung aller Dummy-Punkte  $\mathcal{D}_{d,i} = \{\underline{u}_{d,i}(1), \underline{u}_{d,i}(2), \dots, \underline{u}_{d,i}(N_{d,i})\}$  in einem Schritt  $i$  erfolgt mit einem raumfüllenden Versuchsplan in dem zuvor betrachteten Unterraum. Hierzu wird ein LHC verwendet, der mittels EDLS-Algorithmus [26] optimiert wird.

Das Teilmodell in einem Schritt  $i \geq 2$  wird auf Basis aller Punkte  $\mathcal{D}_i \cup \mathcal{D}_{d,i}$  aus dem jeweiligen inkrementellen Versuchsplan  $\mathcal{D}_i$  und den zugehörigen Dummy-Punkten  $\mathcal{D}_{d,i}$  erstellt, sodass sich für das GPM eine  $(N_i + N_{d,i}) \times (N_i + N_{d,i})$  dimensionale Kovarianzmatrix ergibt. Während zu den Punkten in  $\mathcal{D}_i$  eine gemessene Ausgangsgröße vorliegt, für die ein Rauschen mit der Standardabweichung  $\hat{\sigma}_n$  angenommen wird, ist für alle Dummy-Punkte in  $\mathcal{D}_{d,i}$  ein Ausgangswert von exakt null ohne Rauschen gefordert. Aus diesem Grund wird Gleichung (2.15), in der die Rauschannahme in die Kovarianzmatrix einbezogen wird, abgewandelt zu:

$$K(\underline{u}(k), \underline{u}(l)) \rightarrow \begin{cases} K(\underline{u}(k), \underline{u}(l)) + \hat{\sigma}_n^2 & \text{für } k = l \text{ und } \underline{u}(k) \in \mathcal{D}_i \\ K(\underline{u}(k), \underline{u}(l)) & \text{sonst,} \end{cases} \quad (3.15)$$

mit  $k = 1, 2, \dots, N_i + N_{d,i}$  und  $l = 1, 2, \dots, N_i + N_{d,i}$ . Das Rauschen wird nur für Punkte außerhalb des Unterraums angenommen, während das GPM innerhalb des Unterraums interpoliert. Dieses GPM besitzt somit eine von den Eingangsgrößen abhängige Rauschannahme. Für eine Betrachtung von GPM mit komplexen, von den Eingangsgrößen abhängigen Rauschannahmen sei auf [38] verwiesen. Um eine Verfälschung der Hyperparameter und numerische Probleme durch die Dummy-Punkte zu vermeiden, wird zunächst ein GPM ohne Berücksichtigung der Dummy-Punkte auf Basis von  $\mathcal{D}_i$  erstellt. Diese Hyperparameter werden anschließend für das Teilmodell, das auch die Dummy-Punkte mit einschließt, übernommen.

Bild 3.17a zeigt den Modellausgang des Teilmodells  $f_2$  mit den verwendeten Trainingsdatenpunkten und Dummy-Punkten. Der Modellausgang in dem Unterraum mit  $u_2 = \bar{u}_2 = 0.7$  ist gesondert in Bild 3.17b dargestellt, wobei zusätzlich das 95 %-Konfidenzintervall zu sehen ist. Anhand des Konfidenzintervalls wird deutlich, dass die hier beschriebene Methode GPM-DP in dem zuvor betrachteten Unterraum keinen exakten Modellausgang von null erzeugt, sondern sehr kleine Abweichungen möglich



(a) Teilmodellausgang von  $f_2$  mit zugehörigen Trainings- und Dummy-Punkten

(b) Modellausgang und 95 % Konfidenzintervall von  $f_2$  im Unterraum  $u_2 = \bar{u}_2$

Bild 3.17: Das GPM-DP Teilmodell  $f_2$  wird in (a) auf Basis der Trainings- und Dummy-Punkte erstellt. Der Modellausgang im Unterraum (graue Ebene) ergibt sich in (b) durch die Dummy-Punkte näherungsweise zu null [114].

sind. Dies liegt zum einen an einem zusätzlichen Offset (*Ridge Regularisierung*) von  $10^{-8}$ , der aus numerischen Gründen zu allen Diagonalelementen der Kovarianzmatrix addiert wird und somit kleine Abweichungen an den Dummy-Punkten zulässt. Zum anderen ist das Modellverhalten zwischen den Dummy-Punkten nicht spezifiziert und kann von null abweichen. Die auftretenden Abweichungen sind unter anderem von der Anzahl an verwendeten Dummy-Punkten abhängig, für die im Folgenden dieselbe Anzahl  $N_{d,i} = N_i$  gewählt wird, wie Trainingsdatenpunkte vorhanden sind.

### GPM mit nicht stationärer Kernel-Funktion

Typische Kernel-Funktionen wie ein Gaußkern gemäß Gleichung (2.17) sind stationär. Die Charakteristik dieser Kernel-Funktionen ist unabhängig von den Eingangsgrößen  $\underline{u}$  und damit von der Position im Eingangsraum [88]. Soll durch eine Kernel-Funktion ein Modellausgang von null im zuvor betrachteten Unterraum erzeugt werden, muss sich die Charakteristik der Kernel-Funktionen in Abhängigkeit der Eingangsgrößen  $\underline{u}$  ändern, was als nicht stationäres Verhalten bezeichnet wird. Der hier gewählte Ansatz basiert auf der Interpretation des Modellausgangs eines GPM als gewichtete Summe der auf jedem Trainingsdatenpunkt platzierten Kernel-Funktionen (vgl. Gleichung (2.19)). Demnach ist an einer Position im Eingangsraum

ein Modellausgang von null gegeben, wenn alle Kernel-Funktionen an dieser Position einen Wert von null aufweisen. Nachfolgend wird eine stationäre Kernel-Funktion, die für zwei Punkte  $\underline{u}(k)$  mit  $k = 1, 2, \dots, N_i$  und  $\underline{u}(l)$  mit  $l = 1, 2, \dots, N_i$  ausgewertet wird, mithilfe zweier Linearfaktoren modifiziert:

$$K(\underline{u}(k), \underline{u}(l)) \rightarrow (u_i(k) - \bar{u}_i) \cdot (u_i(l) - \bar{u}_i) \cdot K(\underline{u}(k), \underline{u}(l)). \quad (3.16)$$

Diese Linearfaktoren werden zu null, sobald die  $i$ -te Koordinate  $u_i$  von einem der betrachteten Punkte  $\underline{u}(k)$  oder  $\underline{u}(l)$  den Wert des Arbeitspunkts  $\bar{u}_i$  aufweist und somit einer der Punkte im zuvor betrachteten Unterraum liegt. Nachfolgend wird diese Modifikation auf einen Gaußkern angewandt. Wegen der resultierenden nicht stationären (NS) Kernel-Funktion wird diese Methode als GPM-NS bezeichnet.

GPM-NS verwendet für das Training eines Teilmodells ausschließlich die  $N_i$  Datenpunkte des inkrementellen Versuchsplans, während bei GPM-DP die zusätzlichen Dummy-Punkte zu insgesamt  $N_i + N_{d,i}$  Trainingsdatenpunkten führen. Die resultierende  $N_i \times N_i$  Kovarianzmatrix führt bei GPM-NS zu einem verringerten Rechenaufwand gegenüber der  $(N_i + N_{d,i}) \times (N_i + N_{d,i})$  Kovarianzmatrix bei GPM-DP. Zusätzlich ist ein exakter Modellausgang von null im gesamten zuvor betrachteten Unterraum sichergestellt. Allerdings folgt die Modifikation der Kernel-Funktion in Gleichung (3.16) nicht den Konstruktionsprinzipien für Kernel-Funktionen, wie sie in [36] beschrieben sind. Die resultierende Funktion erfüllt damit nicht die Anforderung einer positiv definiten Funktion, sodass die mathematischen Bedingungen für zulässige Kernel-Funktionen nicht gegeben sind. Jedoch kann dieser Ansatz als regularisiertes *radiales Basisfunktionsnetz* interpretiert werden, bei dem auch Basisfunktionen zulässig sind, die keine positive Definitheit aufweisen [70]. Eine zuverlässige Angabe eines Konfidenzintervalls ist somit allerdings nicht möglich, weil die hierfür notwendige Bayes'sche Interpretation des Modells nur bei positiv definiten Kernel-Funktionen gültig ist.

### Interpretation der Kernel-Standardabweichungen

GPM ermögliche es, irrelevante Eingangsgrößen anhand der Standardabweichungen der Kernels zu detektieren. Hierfür werden Kernel-Funktionen wie der Gaußkern in Gleichung (2.17) mit separaten Standardabweichungen für jede Dimension benötigt, die als ARD-Kernels (engl. *automatic relevance determination*, kurz: ARD) bezeichnet werden [88]. Die Hyperparameteroptimierung während des Trainings eines

GPM führt dann dazu, dass irrelevanten Eingangsgrößen große Standardabweichungen zugewiesen werden, wodurch der Modellausgang nahezu unabhängig von diesen Eingangsgrößen ist.

In der inkrementellen Modellstruktur können mit diesem Mechanismus Haupteffekte und Interaktionen der einzelnen Eingangsgrößen identifiziert werden, indem die Kernel-Standardabweichungen der Teilmodelle verglichen werden. Tabelle 3.2 stellt die Kernel-Standardabweichungen  $\sigma_j$  mit  $j = 1, 2, \dots, 5$  der Teilmodelle  $f_j$  und des Vergleichsmodells dar, wobei die Testfunktion Friedman gemäß Gleichung (3.4) modelliert wird. Alle sehr großen Standardabweichungen z. B. über einem Schwellwert von  $\sigma_j > 1000$  deuten auf eine für das Teilmodell irrelevante Eingangsgröße hin und sind grau hinterlegt.

Jedes Teilmodell  $f_j$  in der inkrementellen Struktur bezieht eine weitere Eingangsgröße  $u_j$  mit der zugehörigen Kernel-Standardabweichung  $\sigma_j$  in das Gesamtmodell ein und bildet so den Haupteffekt dieser Eingangsgröße ab. Aus den kleinen Werten dieser Standardabweichungen (blaue Markierungen) kann geschlossen werden, dass alle Eingangsgrößen einen signifikanten Haupteffekt aufweisen. Im Teilmodell  $f_3$  ist auch die zuvor einbezogene Eingangsgröße  $u_2$  mit einer ebenfalls kleinen Standardabweichung  $\sigma_2 \approx \sigma_3$  versehen (orange Markierung), was auf Interaktionen zwischen den Eingangsgrößen  $u_2$  und  $u_3$  hindeutet. Für alle anderen Teilmodelle  $f_j$  mit  $j = 2, 4, 5$  sind mit  $\sigma_l \gg \sigma_j \forall l = 1, 2, \dots, j - 1$  alle zuvor betrachteten Eingangsgrößen irrelevant, sodass keine weiteren Interaktionen erkennbar sind. Auf diese Weise spiegelt sich die Struktur der verwendeten Testfunktion mit Interaktionen zwischen  $u_2$  und  $u_3$  und den unabhängigen Eingangsgrößen  $u_1, u_4$  und  $u_5$  korrekt in den Kernel-Stan-

Tabelle 3.2: Vergleich der Kernel-Standardabweichungen zwischen einem inkrementellen Modell am Beispiel von GPM-DP und dem Vergleichsmodell bei der Modellierung der Testfunktion Friedman gemäß Gleichung (3.4).

Modell- typ	Teil- modell	Kernel-Standardabweichungen				
		$\sigma_1$	$\sigma_2$	$\sigma_3$	$\sigma_4$	$\sigma_5$
GPM- DP	$f_1$	3.4				
	$f_2$	$2.6 \cdot 10^6$	1.1			
	$f_3$	$4.1 \cdot 10^6$	0.7	0.7		
	$f_4$	$1.2 \cdot 10^7$	$1.7 \cdot 10^7$	$3.1 \cdot 10^{17}$	2.5	
	$f_5$	$1.2 \cdot 10^{15}$	$1.1 \cdot 10^8$	$1.6 \cdot 10^6$	$4.5 \cdot 10^6$	1.0
Vergleichs- modell		4.7	1.4	1.5	43.0	89.6

dardabweichungen wider. Im Gegensatz dazu verdeutlichen die Kernel-Standardabweichungen des Vergleichsmodells mit Werten unterhalb des gewählten Schwellwerts ausschließlich eine Relevanz aller Eingangsgrößen. Aus einem inkrementellen GPM können somit zusätzliche Informationen über die abgebildeten funktionalen Zusammenhänge extrahiert werden.

### Demonstration an einem Testprozess

Um Unterschiede zwischen GPM-DP und GPM-NS herauszustellen, erfolgt eine Modellbildung am Beispiel der Testfunktion Friedman gemäß Gleichung (3.4) in  $p = 3$  Schritten und mit einem erhöhten Rauschlevel von  $\sigma_n = 10^{-3}$ . Bild 3.18 stellt den RMSE dieser Modelle und des Vergleichsmodells für (a) eine Testdatenverteilung mit  $\sigma_{\text{test}} = 0.05$  und (b) eine Sobol-Sequenz gegenüber. Alle Modelle wurden auf Basis einer Gesamtzahl von  $N_\Sigma = 91$  Versuchspunkten erstellt, die bei den inkrementellen Modellen unterschiedlich auf die einzelnen Schritte verteilt sind. Neben einer konstanten Anzahl an Punkten (blau) ist ein linearer Anstieg der Anzahl an Punkten (rot) mit jedem Schritt dargestellt. Der lineare Anstieg führt im gewählten Beispiel bei beiden inkrementellen GPM und beiden Testdatensätzen zu einem verringerten Median des RMSE.

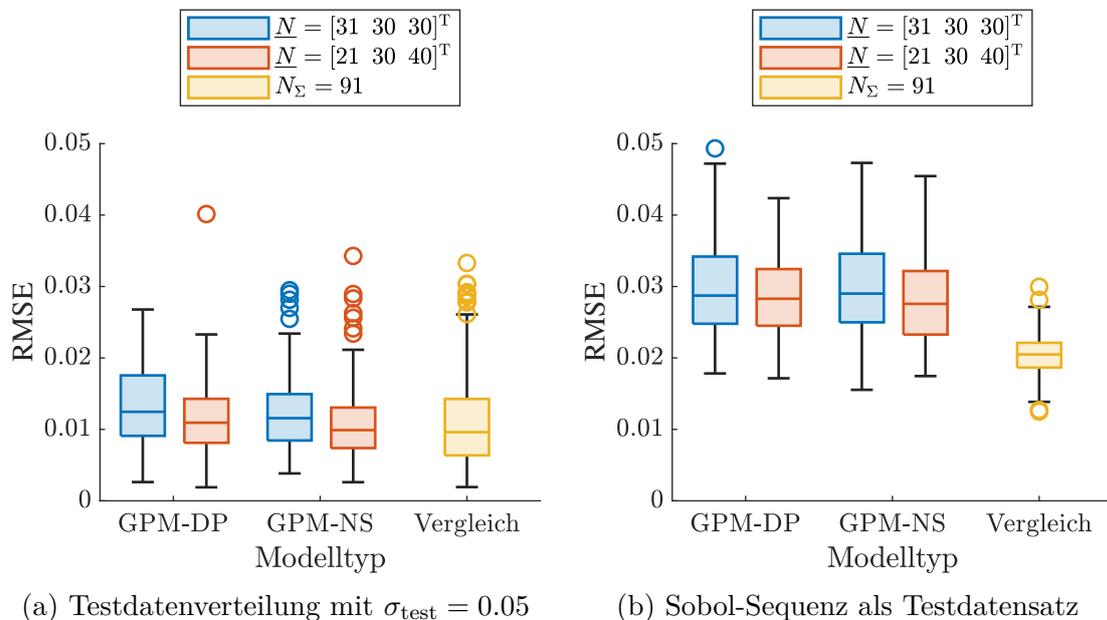


Bild 3.18: Vergleich verschiedener GPM anhand der Testfunktion Friedman gemäß Gleichung (3.4). Der RMSE der Modelle ist für zwei Testdatenverteilungen über 100 Wiederholungen dargestellt.

Insgesamt kann mit beiden Typen an inkrementellen GPM in der Nähe des Arbeitspunkts in (a) eine vergleichbare Güte wie mit dem Vergleichsmodell erzielt werden. Wird hingegen in (b) der gesamte Eingangsraum betrachtet, ist das Vergleichsmodell aufgrund der gleichmäßigeren Datenverteilung im Vorteil. Auch die kleinere Streuung des RMSE beim Vergleichsmodell in (b) ist auf die gleichmäßigere Verteilung der Versuchspunkte zurückzuführen, da hierdurch zufällige Einflüsse durch Rauschen oder die Initialisierung der LHC einen geringeren Einfluss haben.

Im Vergleich der inkrementellen Modelle erzielt GPM-NS einen etwas geringeren mittleren Fehler als GPM-DP. Dies kann durch die unmittelbare Berücksichtigung des geforderten Verhaltens im zuvor betrachteten Unterraum bei der Hyperparameteroptimierung erklärt werden, während das Modellverhalten bei GPM-DP nachträglich durch die zusätzlichen Dummy-Punkte verändert wird. Im Gegenzug ist GPM-NS durch die durchgeführte Modifikation der Kernel-Funktion stärker von numerischen Fehlern betroffen als GPM-DP, was zu einem RMSE mit einer größeren Streuung und mehr Ausreißern führt.

## 3.4 Beurteilung der Methodik

In diesem Abschnitt werden Vor- und Nachteile der ILHAD-Methodik diskutiert. Diese Vor- und Nachteile beziehen sich sowohl allgemein auf die Methodik als auch auf einzelne Modellklassen. Neben der erzielbaren Modellgüte stehen weitere Aspekte wie die Einbeziehung von Prozesswissen und die praktische Anwendbarkeit im Fokus.

### 3.4.1 Bewertung der Modellgüte und -komplexität

Die verschiedenen Modellklassen in der inkrementellen Modellbildung wurden in den vorherigen Abschnitten betrachtet und einzeln untersucht. In diesem Abschnitt erfolgen Bewertung und Vergleich der entworfenen Teilmodellstrukturen miteinander. Alle inkrementellen Modelle werden mit  $\underline{N} = [21 \ 40 \ 60]^T$  Trainingsdatenpunkten und die zugehörigen Vergleichsmodelle mit in Summe derselben Anzahl  $N_{\Sigma} = 121$  an Trainingsdatenpunkten in  $p = 3$  Schritten erstellt, wobei die Testfunktion Gauß gemäß Gleichung (3.3) und die Testfunktion Friedman gemäß Gleichung (3.4) modelliert werden. Alle Modelle werden mit den Standardparametern aus Abschnitt 3.3

erstellt. Für das inkrementelle Modell mit LMN wird mit  $r = 1$  keine Transformation der LHC-Stufen durchgeführt, ein Linearfaktor als Nullstellenfunktion eingesetzt und eine Fehlerschwelle von  $\epsilon = 0.005$  als Abbruchkriterium für den LOLIMOT-Algorithmus verwendet.

### Vergleich der Modellgüte

Tabelle 3.3 fasst den RMSE der verschiedenen Modelle auf Basis von zwei Testdatenverteilungen zusammen. Anhand der Testdaten mit  $\sigma_{\text{test}} = 0.05$  wird deutlich, dass die inkrementellen Modelle typischerweise eine höhere Güte in einem Bereich um den Arbeitspunkt erzielen als das jeweilige Vergleichsmodell. Wird hingegen mit der Sobol-Sequenz der gesamte Eingangsraum betrachtet, erzielen die Vergleichsmodelle aufgrund der gleichmäßigeren Datenverteilung bessere Ergebnisse. Eine Ausnahme stellen die inkrementellen Modelle mit LMN dar, die von der geringen Parameterzahl der lokalen Modelle in den ersten Schritten profitieren. Dies ermöglicht gegenüber den Vergleichsmodellen eine insgesamt größere Anzahl an lokalen Modellen, was auch

Tabelle 3.3: Vergleich der Modellklassen anhand zweier Testfunktionen. Für jedes Modell sind mit  $\mu \pm \sigma$  Mittelwert  $\mu$  und Standardabweichung  $\sigma$  des RMSE auf Basis zweier Testdatenverteilungen angegeben.

Test-funktion	Modell-klasse	Art	RMSE $\cdot 10^3$ $\sigma_{\text{test}} = 0.05$	RMSE $\cdot 10^3$ Sobol	
Gauß (3.3)	Polynom	Inkrementell	$26.84 \pm 0.75$	$53.82 \pm 1.16$	
		Vergleichsmodell	$27.56 \pm 0.58$	$48.63 \pm 0.01$	
	LMN	Inkrementell	$13.62 \pm 0.74$	$22.54 \pm 4.14$	
		Vergleichsmodell	$49.29 \pm 0.50$	$29.56 \pm 0.04$	
	GPM	GPM-DP GPM-NS Vergleichsmodell		$0.64 \pm 0.26$	$2.66 \pm 0.64$
				$0.85 \pm 0.99$	$4.10 \pm 4.58$
			$0.57 \pm 0.29$	$0.96 \pm 0.11$	
Friedman (3.4)	Polynom	Inkrementell	$11.22 \pm 1.86$	$104.21 \pm 0.98$	
		Vergleichsmodell	$54.62 \pm 0.36$	$89.76 \pm 0.01$	
	LMN	Inkrementell	$6.82 \pm 0.45$	$15.88 \pm 2.60$	
		Vergleichsmodell	$21.90 \pm 0.43$	$33.34 \pm 0.05$	
	GPM	GPM-DP GPM-NS Vergleichsmodell		$0.44 \pm 0.20$	$1.51 \pm 0.51$
				$0.47 \pm 0.19$	$1.69 \pm 0.53$
			$0.72 \pm 0.29$	$1.26 \pm 0.12$	

für den gesamten Eingangsraum zu einer genaueren Abbildung des Prozessverhaltens führt. Werden die Fehler der einzelnen Modellklassen miteinander verglichen, ist eine Verbesserung mit zunehmender Flexibilität der Modelle erkennbar. Die Polynommodelle mit der geringsten Flexibilität schneiden insgesamt am schlechtesten ab, während die GPM mit der höchsten Flexibilität zu den besten Ergebnissen führen.

Um das Verhalten der inkrementellen Modellbildung bei den verschiedenen Testfunktionen zu vergleichen, werden die GPM betrachtet, da diese auf geringen Annahmen über den betrachteten Prozess beruhen. Bei den Polynommodellen und LMN können hingegen der polynomiale Grad oder die Abbruchbedingung im LOLIMOT-Algorithmus unterschiedlich gut für die jeweilige Testfunktion geeignet sein und so die Ergebnisse beeinflussen. Die inkrementellen GPM sind für  $\sigma_{\text{test}} = 0.05$  bei der Testfunktion Friedman im Vorteil, während das Vergleichsmodell bei der Gaußfunktion die besten Ergebnisse erzielt. Für die Sobol-Sequenz als Testdatenverteilung sind bei der Testfunktion Friedman die Mittelwerte des RMSE der inkrementellen Modelle gegenüber dem Vergleichsmodell um 20 % bei GPM-DP und 34 % bei GPM-NS erhöht. Bei der Gaußfunktion steigen diese Fehler um 277 % bei GPM-DP und 427 % bei GPM-NS gegenüber dem Vergleichsmodell an. Insgesamt zeigt sich somit, dass die inkrementellen Modelle besser in der Lage sind, die funktionalen Zusammenhänge der Testfunktion Friedman abzubilden, als die der Gaußfunktion.

Der größere Nutzen der inkrementellen Modellbildung bei der Testfunktion Friedman kann durch die vorliegenden Interaktionen der Eingangsgrößen in Verbindung mit dem verwendeten Versuchsplan erklärt werden. Wird in einem Schritt  $i$  eine Eingangsgröße  $u_i$  in ein inkrementelles Modell einbezogen, sind alle zuvor betrachteten Eingangsgrößen, die nicht mit  $u_i$  interagieren, irrelevant für das jeweilige Teilmodell. Relevante Informationen sind somit nur in einer Projektion des verwendeten Versuchsplans  $\mathcal{D}_i$  ohne diese irrelevanten Eingangsgrößen enthalten. Durch die LHC-Eigenschaft von  $\mathcal{D}_i$  weist jede mögliche Projektion gute raumfüllende Eigenschaften auf (siehe Abschnitt 2.4.2) und ermöglicht auf diese Weise eine hohe Modellgüte. Bei der Testfunktion Friedman kann diese LHC-Eigenschaft in jedem Schritt genutzt werden, da weder  $u_2$  noch  $u_3$  mit der Eingangsgröße  $u_1$  interagieren. Bei der Testfunktion Gauß kann dieser Vorteil von ILHAD nicht genutzt werden, da alle Eingangsgrößen miteinander interagieren.

Für alle Modellklassen sind in Tabelle 3.3 typischerweise größere Standardabweichungen des RMSE für die inkrementellen Modelle als für die Vergleichsmodelle gegeben. Dies verdeutlicht die größere Sensitivität der inkrementellen Modelle auf zufällige

Einflüsse wie Prozessrauschen oder die Initialisierung der LHC in jedem Schritt des inkrementellen Versuchsplans. Durch das Konstruktionsprinzip der inkrementellen Modelle sind diese Schwankungen des RMSE bei Betrachtung des gesamten Eingangsraums mit der Sobol-Sequenz größer als in der Nähe des Arbeitspunkts mit  $\sigma_{\text{test}} = 0.05$ . Für eine Untersuchung der Auswirkungen von Prozessrauschen auf die inkrementelle Modellbildung sei auf [114] verwiesen.

### Vergleich der Modellkomplexität

Neben dem Modellfehler ist die Komplexität der einzelnen Modelle von Bedeutung. Von mehreren Modellen mit vergleichbarer Güte ist typischerweise dasjenige zu bevorzugen, das die geringste Komplexität aufweist. Dieses Prinzip wird als *Ockhams Rasiermesser* bezeichnet und führt zu einer höheren Interpretierbarkeit, einer größeren Robustheit und einem geringeren Speicherbedarf der Modelle [23]. Tabelle 3.4 stellt die Komplexität der verschiedenen Modelle anhand der Anzahl an lokalen Modellen und Parametern gegenüber. Hierbei sind exemplarisch die Modelle aus Tabelle 3.3 aufgeführt, die zur Modellierung der Gaußfunktion dienen.

Für das inkrementelle Polynommodell und GPM-NS ist jeweils dieselbe Anzahl an Modellparametern gegeben wie für das jeweilige Vergleichsmodell. Bei GPM-DP ist die Anzahl an Modellparametern durch die zusätzlichen Dummy-Punkte in den Schritten  $i \geq 2$  erhöht. Zu beachten ist, dass die effektive Anzahl an Parametern bei allen GPM durch die Regularisierung verringert wird und so von den jeweiligen Hyperparametern des GPM abhängt [80]. Bei einem inkrementellen Modell mit

Tabelle 3.4: Anzahlen an lokalen Modellen und Parametern der in Tabelle 3.3 zur Modellierung der Gaußfunktion verwendeten Modelle.

Modellklasse	Art	Anzahl lokaler Modelle	Anzahl Parameter
Polynom	Inkrementell	/	10
	Vergleichsmodell	/	10
LMN	Inkrementell	22	72
	Vergleichsmodell	19	76
GPM	GPM-DP	/	221
	GPM-NS	/	121
	Vergleichsmodell	/	121

LMN besitzt ein lokal affines Modell, das in einem Schritt  $i$  erstellt wird,  $i + 1$  Parameter. Ein lokal affines Modell des Vergleichsmodells besitzt hingegen immer  $p + 1$  Parameter. Bei gleicher Anzahl an lokalen Modellen sind somit insgesamt weniger effektive Parameter in einem inkrementellen Modell enthalten. Dies ermöglicht dem inkrementellen Modell mit einer gegebenen Anzahl an Versuchspunkten mehr lokale Modelle zu erstellen als das Vergleichsmodell.

### 3.4.2 Berücksichtigung von Prozesswissen

In der praktischen Durchführung von Versuchen müssen unweigerlich Annahmen über den betrachteten Prozess getroffen werden, die Auswirkungen auf die Versuchsdurchführung und die Modellbildung haben. Das hierzu eingesetzte Prozesswissen trägt dazu bei, erfolgreiche Versuche durchzuführen und eine hohe Modellqualität zu erzielen [77]. Folgende Aufgaben der experimentellen Modellbildung sind in besonderem Maße von diesem Wissen abhängig:

1. Auswahl der im Versuch betrachteten Eingangsgrößen,
2. Definition der zulässigen Bereiche im Eingangsraum,
3. Auswahl eines Versuchsplans bzw. einer Modellstruktur (z. B. bei optimalen Versuchsplänen).

Durch das schrittweise Vorgehen der ILHAD-Methodik müssen diese Aufgaben nicht im Vorfeld der Versuche abgeschlossen sein, sondern können für jeden einzelnen Schritt bearbeitet und angepasst werden. Hierbei kann auf Wissen zurückgegriffen werden, das bereits in den vorherigen Schritten über den Prozess erlangt wurde. Somit ergeben sich folgende Anpassungen der zuvor aufgeführten Aufgaben:

1. Anstatt alle betrachteten Eingangsgrößen auszuwählen, wird vor Beginn der Versuche eine Rangfolge der Eingangsgrößen gemäß ihrer Bedeutung für den Prozess festgelegt. Mit einer sinnvoll gewählten Rangfolge kann bei vielen Prozessen bereits in wenigen Schritten eine hohe Modellgüte erzielt werden. Die Sicherheit, mit der die Bedeutung einer Eingangsgröße beurteilt werden kann, kann zusätzlich in die Rangfolge einfließen. Eingangsgrößen, bei denen eine große Unsicherheit über die Bedeutung besteht, können in der Rangfolge an hintere Positionen verschoben werden. Dies ermöglicht, diese Größen vorerst unberücksichtigt zu lassen und nur bei Bedarf in das inkrementelle Modell aufzunehmen. Bei einer Vorauswahl müssten solche Größen direkt im Versuchsplan

berücksichtigt werden, um keine potentiell wesentlichen Einflüsse zu vernachlässigen, was die Dimensionalität und den Versuchsaufwand steigert.

2. Viele Prozesse weisen Grenzwerte der Ausgangsgrößen auf, die nicht unter- oder überschritten werden dürfen. Für einen Versuchsplan muss daher der zulässige Eingangsraum definiert werden. Die Komplexität dieser Aufgabe nimmt mit steigender Dimensionalität des Eingangsraums zu, da die überlagerten Auswirkungen einer zunehmenden Anzahl an Eingangsgrößen beurteilt werden müssen. Die ILHAD-Methodik vereinfacht die Festlegung von zulässigen Bereichen durch die niedrige Dimensionalität in den ersten Schritten. Zusätzlich kann in den Schritten  $i \geq 2$  auf Wissen aus vorherigen Schritten zurückgegriffen werden, das bereits durch das inkrementelle Modell abgebildet ist. Kritische Bereiche im Eingangsraum, in denen die Ausgangsgröße Werte in der Nähe vorgegebener Grenzwerte aufweisen, können in einem Schritt  $i$  anhand des aktuellen Modells bestimmt werden. Im nächsten Schritt  $i + 1$  müssen hauptsächlich die Auswirkungen der zusätzlichen Eingangsgröße  $u_{i+1}$  in diesen Bereichen beurteilt werden, was die Definition des zulässigen Eingangsraums erleichtert.
3. Der inkrementelle Versuchsplan ist durch seine in jedem Schritt raumfüllende Eigenschaft für verschiedene Modellklassen geeignet und ermöglicht es, verschiedene inkrementelle Modelle zu erstellen und zu vergleichen. Auf diese Weise können die Vorteile der einzelnen Modellklassen wie die einfache Interpretierbarkeit der Polynommodelle, die Detektion von Nichtlinearitäten mithilfe der durch LOLIMOT erstellten LMN-Partitionen und die Unterscheidung von Haupteffekten und Interaktionen zwischen den Eingangsgrößen anhand der Kernel-Standardabweichungen des GPM kombiniert werden.

Eine zusätzliche Aufgabe in der ILHAD-Methodik, die auf der Grundlage von Prozesswissen erfolgen muss, ist die Festlegung des Arbeitspunkts. Durch das Konstruktionsprinzip des inkrementellen Versuchsplans und die Teilmodellstrukturen kann eine hohe Güte eines inkrementellen Modells in der Nähe dieses Arbeitspunkts erwartet werden. Trotzdem wird durch die Wahl des Arbeitspunkts auch die Modellqualität im gesamten Eingangsraum beeinflusst, wenn Interaktionen zwischen den Eingangsgrößen vorliegen. Je nach Wahl des Arbeitspunkts können mit einem inkrementellen Modell funktionale Zusammenhänge mit unterschiedlicher Komplexität erfasst und abgebildet werden. Bild 3.19 stellt solche unterschiedlich komplexen funktionalen Zusammenhänge am Beispiel der Testfunktion Friedman gemäß Gleichung (3.4) dar, wobei die Koordinate  $\bar{u}_3$  des Arbeitspunkts variiert wird. Bei der Wahl des Arbeitspunkts sollten daher zwei Ziele abgewogen werden. Zum einen sollte der Arbeitspunkt

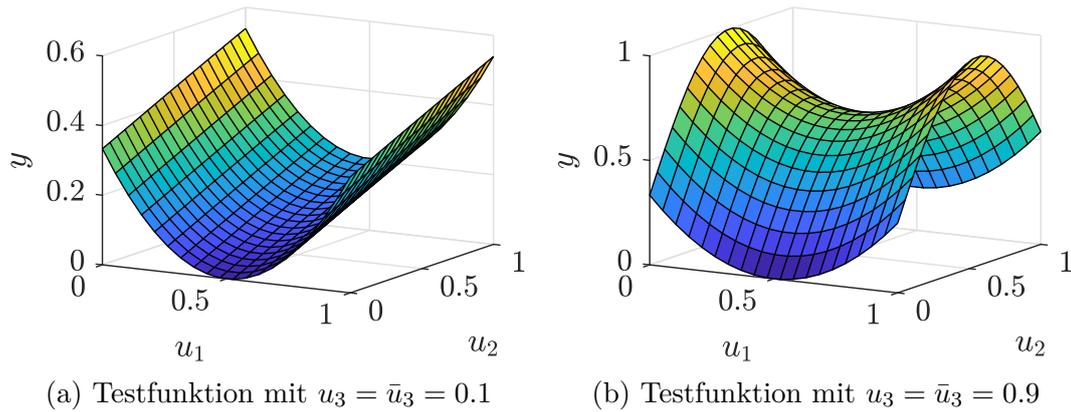


Bild 3.19: Testfunktion Friedman für verschiedene Werte der Koordinate des Arbeitspunkts  $\bar{u}_3$ . Hieraus ergeben sich funktionalen Zusammenhänge mit unterschiedlicher Komplexität, die approximiert werden müssen.

aufgrund der in dessen Nähe zu erwartenden hohen Güte möglichst repräsentativ für die spätere Verwendung des Modells sein. Zum anderen sollte er so gewählt werden, dass jedes Teilmodell funktionale Zusammenhänge mit einer angemessenen Komplexität abbilden kann.

### 3.4.3 Praktische Anwendbarkeit

Eine erfolgreiche experimentelle Modellbildung hängt wie in Abschnitt 3.4.2 beschrieben in hohem Maße von dem verfügbaren Prozesswissen ab. Die in diesem Kapitel vorgestellte ILHAD-Methodik bietet einen intuitiven Zugang, um dieses Prozesswissen in den Modellbildungsprozess einfließen zu lassen. Hierzu trägt die geringe Dimensionalität des Eingangsraums in den ersten Schritten bei. Insbesondere sind die Versuchspläne für  $i \leq 3$  und das Modellverhalten für  $i \leq 2$  einfach visualisierbar. Die ILHAD-Methodik ermöglicht somit in den ersten Schritten ein Verständnis des grundlegenden Prozessverhaltens. Zwar existieren zahlreiche Methoden zur Visualisierung hochdimensionaler Daten [69], allerdings sind diese Darstellungen schwer zu interpretieren und führen stets zu einem Informationsverlust.

#### Eigenschaften zur Erzielung eines geringen Modellierungsaufwands

Die Anzahl an Schritten muss bei ILHAD nicht im Vorfeld der Versuche festgelegt werden. Dies ermöglicht es, die Eignung eines inkrementellen Modells für die praktische Anwendung nach jedem Schritt zu evaluieren. Indem nur so viele Schritte

durchgeführt und damit Eingangsgrößen betrachtet werden, wie notwendig, wird eine geringe Dimensionalität des Eingangsraums erreicht. Dies führt zu einer geringen Komplexität des inkrementellen Modells und trägt zur Interpretierbarkeit bei.

Eine bedeutende Aufgabe bei der Durchführung von Versuchen ist die Einstellung der Eingangsgrößen. Schwer veränderbare Eingangsgrößen, wie beispielsweise eine Temperatur, die sich mit einer großen Zeitkonstante ändert, oder Eigenschaften eines Werkzeugs, das manuell getauscht werden muss, können einen hohen Stellaufwand verursachen. ILHAD ermöglicht es, diesen Stellaufwand gegenüber klassischen Versuchsplänen zu reduzieren, indem solche Eingangsgrößen erst in späteren Schritten in das Modell einbezogen werden. In den vorherigen Schritten, in denen diese Größen auf dem Wert eines Arbeitspunkts konstant gehalten werden, ist nur eine einmalige Einstellung notwendig.

In der inkrementellen Modellstruktur werden die einzelnen Teilmodelle nacheinander erstellt und bleiben trotz der Erweiterung durch zusätzliche Teilmodelle unverändert. Dies ermöglicht einen geringen Rechenaufwand bei der Modellerstellung, da jeweils nur wenig Modellparameter bestimmt werden müssen. Damit ist diese Methode auch für große Versuchsumfänge geeignet, wenn beispielsweise vollautomatische Messungen durchgeführt werden können.

### **Einschränkungen der Anwendbarkeit**

Die inkrementelle Modellstruktur kann beim Vorliegen von Grenzen im Eingangsraum, wie sie beispielsweise bei der Modellierung des Bitumenofens in Abschnitt 5.2.1 gegeben sind, zu einer Extrapolation von Teilmodellen führen. Bild 3.20 verdeutlicht dieses Problem für die ersten beiden Schritte der inkrementellen Modellbildung. Durch die gewählte Koordinate des Arbeitspunkts  $\bar{u}_2 = 1$  und die gegebene Grenze im Eingangsraum (blaue Markierung) wird das Teilmodell  $f_1$  nur mit Trainingsdaten aus dem Wertebereich  $u_1 \in [0, 0.5]$  trainiert. Für das Modell  $f_2$  sind für  $u_2 < 1$  auch Werte im Bereich  $u_1 \in (0.5, 1]$  zulässig, in denen das Modell  $f_1$  extrapoliert. Durch die additive Überlagerung aller Teilmodelle in der inkrementellen Modellstruktur (vgl. Bild 3.5) müssen Abweichungen, die aus der Extrapolation des Modells  $f_1$  resultieren, durch das Modell  $f_2$  ausgeglichen werden, was insbesondere bei komplexem Extrapolationsverhalten problematisch ist. Aus diesem Grund sind bei starken Einschränkungen Teilmodellstrukturen wie LMN mit einer Nullstellenfunktion auf

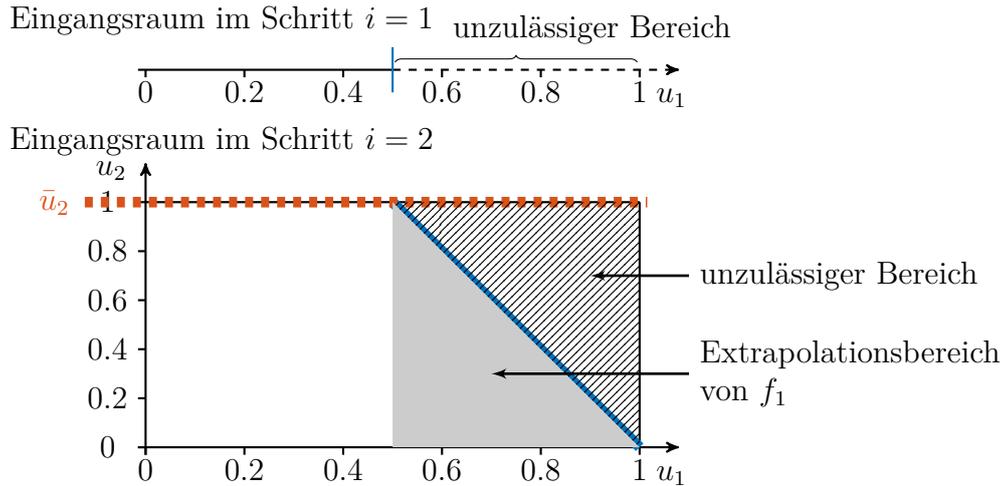


Bild 3.20: Der Eingangsraum im ersten Schritt (oben) ist ein Unterraum (orange Markierung) des Eingangsraums im zweiten Schritt (unten). Die Grenze (blau) führt zusammen mit der Koordinate  $\bar{u}_2 = 1$  des Arbeitspunkts zu einem großen Extrapolationsbereich des Modells  $f_1$  (grau).

Basis von B-Splines zu bevorzugen, die ein lineares Extrapolationsverhalten aufweisen. Auch GPM mit lokalen Kernel-Funktionen sind hierbei vorteilhaft, da deren Modellausgänge bei Extrapolation gegen null streben. Prinzipiell ist eine Extrapolation für alle zuvor erstellten Teilmodelle zu vermeiden. Hierzu sollten die Grenzen bei der Wahl des Arbeitspunkts berücksichtigt werden, sodass insbesondere für die zuerst betrachteten Eingangsgrößen möglichst geringe Einschränkungen vorliegen.

Sind sehr viele Eingangsgrößen hochrelevant, ist ein inkrementelles Modell nur bedingt geeignet. So führt die additive Modellstruktur dazu, dass sich Fehler der einzelnen Teilmodelle insbesondere beim Vorliegen von starkem Prozessrauschen aufsummieren. Eine Korrektur dieser Fehler durch ein nachfolgend erstelltes Teilmodell ist aufgrund der verminderten Punktdichte bei höherer Dimensionalität nicht zielführend. Daher wird eine solche Korrektur durch die geforderte Eigenschaft, dass ein Teilmodell den Modellausgang im zuvor betrachteten Unterraum nicht verändern darf, ausgeschlossen. Die inkrementelle Modellbildung ist aus diesem Grund für Prozesse mit vielen Eingangsgrößen und starkem Prozessrauschen ungeeignet [114].

### 3.5 Zusammenfassung

In diesem Kapitel wurde die *Incremental Latin Hypercube Additive Design* (ILHAD) Methode zur Versuchsplanung und Modellbildung beschrieben, die sich insbesonde-

re für eine Anwendung bei bestehenden mechatronischen Prozessen wie z. B. industriellen Fertigungsanlagen eignet. Das Prozessverhalten wird ausgehend von einem Arbeitspunkt schrittweise erfasst und modelliert, wobei in jedem Schritt eine zusätzliche Eingangsgröße einbezogen wird. Beginnend mit nur einer Eingangsgröße steigt die Dimensionalität des betrachteten Eingangsraums mit jedem Schritt an. Für den jeweiligen Eingangsraum wird mittels eines optimierten Latin Hypercubes (LHC) ein raumfüllender Versuchsplan erstellt. Die auf Basis dieses Versuchsplans erfassten funktionalen Zusammenhänge werden dann durch ein Teilmodell abgebildet und dieses Teilmodell additiv zu einem Gesamtmodell hinzugefügt. Das Gesamtmodell wird somit in jedem Schritt um weitere Informationen ergänzt und verbessert.

Für die Teilmodelle wurden mit Polynommodellen, lokalen Modellnetzen (LMN) und Gaußprozessmodellen (GPM) verschiedene Modellklassen untersucht. Dabei wurden die Teilmodellstrukturen so erweitert, dass das Gesamtmodell im zuvor betrachteten Unterraum des Eingangsraums nicht verändert wird, also das vorherige Modell in seinem Gültigkeitsbereich nicht verfälscht wird. Des Weiteren können Informationen aus den Teilmodellen extrahiert werden, die die folgenden Schritte der inkrementellen Modellbildung unterstützen und eine Validierung des Modells ermöglichen. Es wurde gezeigt, dass Informationen über die Nichtlinearitäten des betrachteten Prozesses zur Verbesserung des inkrementellen Versuchsplans genutzt werden können. Hierzu wurden die Partitions Grenzen der LMN ausgewertet und für eine Transformation der LHC-Stufen im nächsten Schritt verwendet. Für Teilmodelle auf Basis von GPM wurde eine Möglichkeit aufgezeigt, Interaktionen zwischen den Eingangsgrößen anhand der Standardabweichungen der Kernel-Funktion zu detektieren, was zur Steigerung des Prozesswissens beiträgt.

Die Methodik wurde anhand von Simulationen evaluiert und die Ergebnisse anhand eines Vergleichsmodells für jede Modellklasse bewertet. Diese Vergleichsmodelle wurden in einem Schritt auf Basis von optimierten LHC mit derselben Anzahl an Datenpunkten wie in den jeweiligen inkrementellen Versuchsplänen erstellt und repräsentieren ein klassisches Vorgehen zur experimentellen Modellbildung. Es wurde gezeigt, dass inkrementelle Modelle in der Lage sind, eine vergleichbare oder höhere Modellgüte in Bereichen um den Arbeitspunkt zu erzielen wie die jeweiligen Vergleichsmodelle. Dies setzt voraus, dass eine moderate Anzahl an Schritten durchgeführt wird, geringes Prozessrauschen vorliegt und das notwendige Prozesswissen korrekt eingebracht wird. Auf dem Prozesswissen beruhen unter anderem die Auswahl eines Arbeitspunkts, die Definition einer Rangfolge der Eingangsgrößen und die Festlegung der Anzahlen an Versuchspunkten in den einzelnen Schritten.

Insgesamt zielt die inkrementelle Modellbildung darauf ab, den Aufwand in der praktischen Anwendung gegenüber klassischen Verfahren zur experimentellen Modellbildung deutlich zu reduzieren, wobei die Einbußen der Modellgüte möglichst gering ausfallen sollen. Die gute praktische Anwendbarkeit wird durch die schrittweise ansteigende Dimensionalität und damit Komplexität der Versuchsplanung und Modellbildung erzielt. Insbesondere in den ersten Schritten ist auf diese Weise eine einfache Versuchsdurchführung und Bewertung der Ergebnisse z. B. durch eine Visualisierung des Modellausgangs möglich. Da in jedem Schritt ein Prozessmodell vorliegt, das das Verhalten des Prozesses in Abhängigkeit der bisher betrachteten Eingangsgrößen abbildet, kann das Vorgehen jederzeit beendet werden. Auf diese Weise kann mit der inkrementellen Modellbildung bereits in wenigen Schritten ein für die praktische Anwendung geeignetes Modell erstellt werden, das durch eine geringe Dimensionalität einfach zu interpretieren ist und eine geringe Anzahl an Versuchspunkten benötigt.



## 4 Versuchsplanung für zusammengesetzte Einheiten

In vielen realen Prozessen werden Gesamtsysteme aus mehreren *Komponenten* zusammengesetzt. Jede der verwendeten Komponenten wird durch einen festen Satz an Merkmalen beschrieben und beeinflusst die Eigenschaften des Gesamtsystems, sodass aus unterschiedlichen Kombinationen der Komponenten unterschiedliche Gesamtsysteme resultieren. Diese Art von Problemstellung erfordert angepasste Versuchsplanungsmethoden, um in der praktischen Anwendung einen akzeptablen Versuchsaufwand zu erzielen. Hierzu wird in diesem Kapitel ein Vorgehen beschrieben, das mit einem geringen Versuchsaufwand möglichst viele Informationen über einen solchen Prozess erfasst. Im Folgenden werden Montageprozesse als Anwendung für diese Art der Versuchsplanung betrachtet. Ein solcher Montageprozess, in dem zwei unterschiedlich lange Stangen (Komponenten) rechtwinklig zueinander zu *Baugruppen* (Gesamtsysteme) kombiniert werden, ist schematisch in Bild 4.1 dargestellt. Die Versuchsplanungsaufgabe besteht darin, aus allen möglichen Baugruppen diejenigen auszuwählen, die in den Versuchen montiert werden sollen. Die Methodik kann aber auch auf andere Arten von Prozessen angewendet werden, in denen Gesamtsysteme

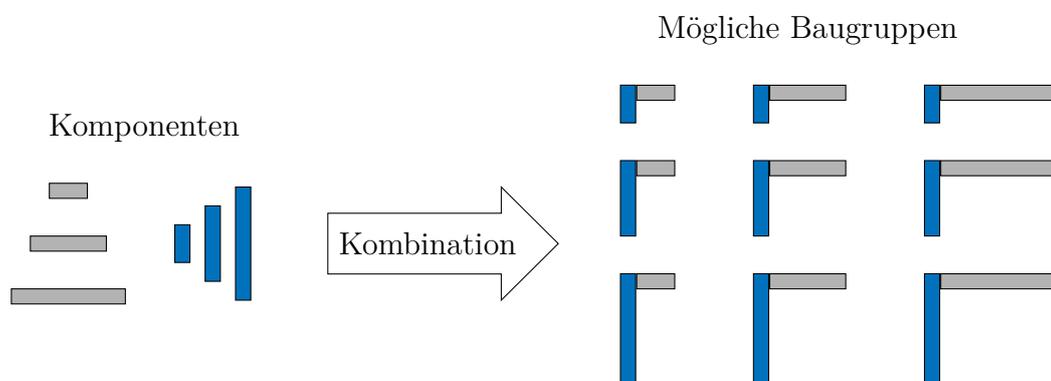


Bild 4.1: Beispiel für einen einfachen Montageprozess, in dem zwei Stangen mit unterschiedlichen Längen rechtwinklig zueinander montiert werden.

aus mehreren Komponenten zusammengesetzt werden. Denkbar wäre z. B. die Kreuzung von Pflanzen oder die Mischung von Vorprodukten, wie sie bei der Entwicklung von Lebensmitteln oder Baustoffen erfolgt.

Das Kapitel ist wie folgt gegliedert: Zunächst wird in Abschnitt 4.1 in das Thema eingeführt und Vorarbeiten aufgezeigt. Die entwickelte Versuchsplanungsmethodik wird in Abschnitt 4.2 beschrieben, wobei ein lokaler und ein globaler Optimierungsalgorithmus verglichen werden. In Abschnitt 4.3 wird ein alternatives Optimierungskriterium betrachtet, mit dem möglichst ähnliche Punkte in einem Versuchsplan erzeugt werden können. Diese ähnlichen Punkte tragen zu einer verbesserten Schätzung der Rauschvarianz bei. Als Maßnahme zur Reduzierung des Versuchsaufwands wird in Abschnitt 4.4 ein Verfahren zur Punktselektion erläutert, das sequentiell die Punkte aus dem Versuchsplan entfernt, von denen der geringste Informationsgewinn erwartet wird. Abschnitt 4.5 beschäftigt sich mit der Montage von rotationssymmetrischen Komponenten und bezieht deren rotatorische Ausrichtung in die Versuchsplanung mit ein. Das Kapitel endet mit der Zusammenfassung der wichtigsten Erkenntnisse in Abschnitt 4.6. Eine reale Anwendung dieses Verfahrens ist am Beispiel einer Trommelmontage für Waschmaschinen in Abschnitt 5.3 zu finden. Teile dieses Kapitels wurden bereits in Form der Konferenzbeiträge [115, 116] veröffentlicht.

## 4.1 Einführung

Montageprozesse werden vielfältig in der Industrie eingesetzt. Hierbei werden Baugruppen aus mehreren Komponenten montiert, wobei verschiedene Verbindungsverfahren wie zum Beispiel Schrauben, Nieten oder Schweißen zum Einsatz kommen können. Die Eigenschaften der resultierenden Baugruppe werden zum einen durch den Montageprozess selbst und die dabei verwendeten Prozessparameter (z. B. Anzugsmoment einer Schraubverbindung oder Schweißparameter), zum anderen durch die Merkmale der verwendeten Komponenten (z. B. Abmessungen oder Materialeigenschaften) beeinflusst. Zur systematischen Untersuchung der Einflüsse von Prozessparametern ist der Einsatz von klassischen Versuchsplanungsmethoden, wie z. B. faktoriellen oder D-optimalen Versuchsplänen üblich [19, 74]. Auch die in Kapitel 3 beschriebene inkrementelle Modellbildung kann hierfür verwendet werden. Für die Untersuchung der Einflüsse der Komponenten sind diese Methoden zumeist nicht geeignet, da eine freie Platzierung von Datenpunkten zumindest auf diskreten Stufen der Merkmale erforderlich ist [77]. Bild 4.2a stellt ein klassisches Vorgehen mit

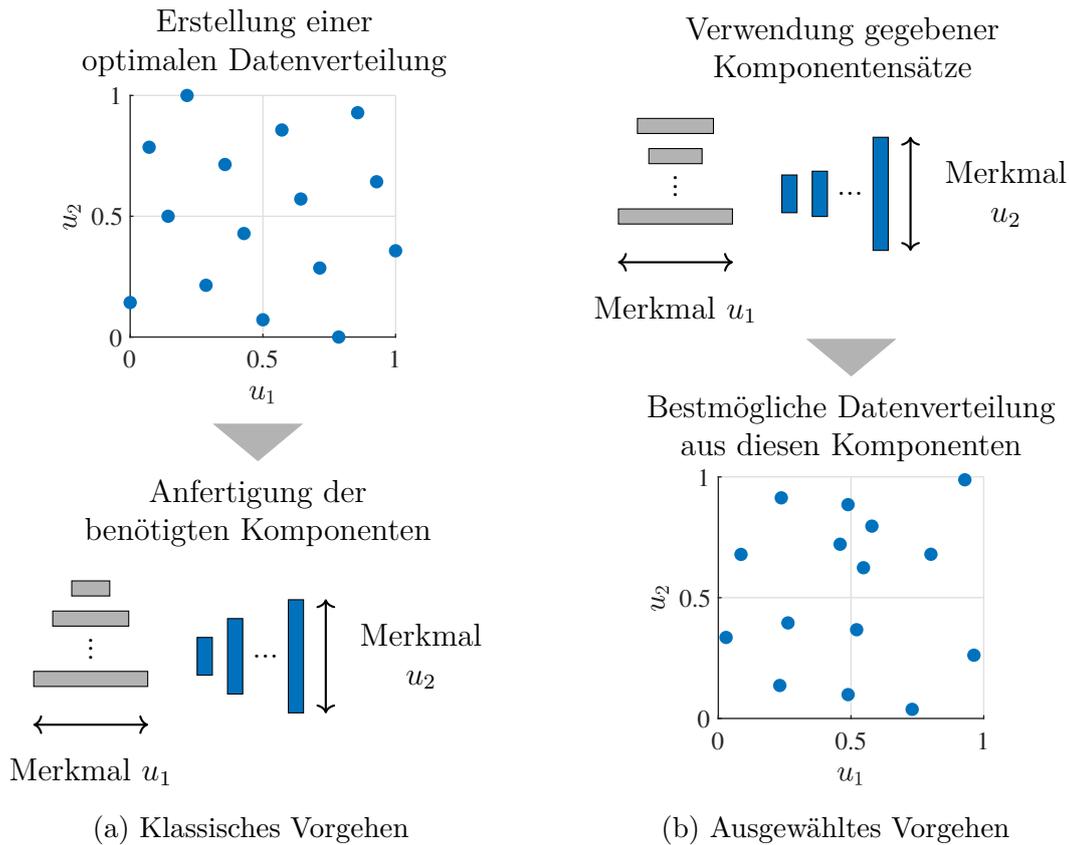


Bild 4.2: Gegenüberstellung eines klassischen Vorgehens zur Versuchsplanung und des in diesen Kapitel gewählten Vorgehens am Beispiel des Montageprozesses aus Bild 4.1.

einem raumfüllenden Versuchsplan dar, in dem die Punkte frei platziert wurden. Experimente mit einem solchen Versuchsplan würden jedoch eine Anfertigung spezieller Komponenten mit den im Versuchsplan festgelegten Merkmalsausprägungen erfordern. Bei vielen Montageprozessen ist eine Anfertigung solcher Komponenten nicht möglich. So können Merkmale nicht in der geforderten Genauigkeit realisiert werden, oder dies würde zu hohen Kosten führen.

Eine Möglichkeit, um kosteneffiziente Versuche an Montageprozessen durchzuführen, ist in Bild 4.2b dargestellt, wobei reguläre Komponenten aus einer laufenden Produktion verwendet werden. Die Komponenten eines jeden Typs bilden einen *Komponentensatz*, in dem jede Einzelkomponente durch die Fertigungssteuerung individuelle Merkmalsausprägungen besitzt. Unterschiedliche Kombinationen dieser Komponenten führen zu unterschiedlichen Datenverteilungen.

Können Baugruppen montiert und wieder demontiert werden, ohne die Eigenschaften der Komponenten zu verändern, ist der Einsatz klassischer Versuchsplanungsverfahren

ren möglich, um verschiedene Kombinationen dieser Komponenten zu untersuchen. In [14] ist ein solches Vorgehen am Beispiel einer Einspritzdüse beschrieben, wobei ein vollfaktorieller Versuchsplan verwendet wird, um mittels statistischer Tests fehlerhafte Komponenten zu identifizieren. Ist eine Demontage nicht ohne Veränderungen an den Komponenten möglich, müssen angepasste Versuchsplanungsmethoden verwendet werden. Eine solche Methode mit einmaliger Verwendung von Komponenten aus gegebenen Sätzen ist in [98] beschrieben und dient zur Erstellung von D-optimalen Versuchsplänen. Hierzu wird ein Austauschalgorithmus verwendet, der mit der *Fedorov-Methode* mögliche Tauschvorgänge hinsichtlich der Verbesserung der D-Optimalität bewertet und denjenigen mit der besten Bewertung durchführt. Zusätzlich liegt ein Schwerpunkt auf der Einbeziehung von Merkmalen, die sich aus geometrischen Beziehungen zwischen mehreren Komponenten ergeben. In [4] ist eine alternative Strategie zur Erstellung von D-optimalen Versuchsplänen für Montageprozesse beschrieben, die einen genetischen Algorithmus verwendet. Hierbei stellt die hohe Anzahl an Hyperparametern eine Herausforderung dar. Beide Verfahren werden in [97] anhand der Montageprozesse einer Hydraulikpumpe und eines Schallwandlers verglichen, wobei der Austauschalgorithmus insgesamt bessere Ergebnisse erzielt.

Ziel der in diesem Kapitel vorgestellten Versuchsplanungsmethodik ist es, die bestmöglichen Kombinationen aus gegebenen Komponentensätzen zu erzeugen, wobei jede Komponente nur einmal verwendet wird. Die resultierenden Datenverteilungen sollen den Eingangsraum möglichst gleichmäßig abdecken. Auf diese Weise soll der funktionale Zusammenhang zwischen den Merkmalen der verwendeten Komponenten und den resultierenden Eigenschaften der Gesamtbaugruppe mit einer möglichst hohen Güte und einem für die industrielle Praxis akzeptablen Versuchsaufwand datenbasiert modelliert werden können.

## 4.2 Versuchsplanungsmethodik

Die nachfolgend beschriebene Versuchsplanungsmethodik zielt darauf ab, mit geringem Aufwand in der praktischen Anwendung möglichst viele Informationen über einen Montageprozess zu erlangen, in dem eine Baugruppe aus  $I$  verschiedenen *Komponententypen* zusammengesetzt wird. Hierzu werden ausschließlich gegebene Sätze an Komponenten verwendet, die jeweils Komponenten eines Typs enthalten. Bild 4.3 zeigt die Zusammenstellung einer Baugruppe aus gegebenen Komponentensätzen. Ein Komponentensatz mit  $N$  Komponenten eines Typs wird als

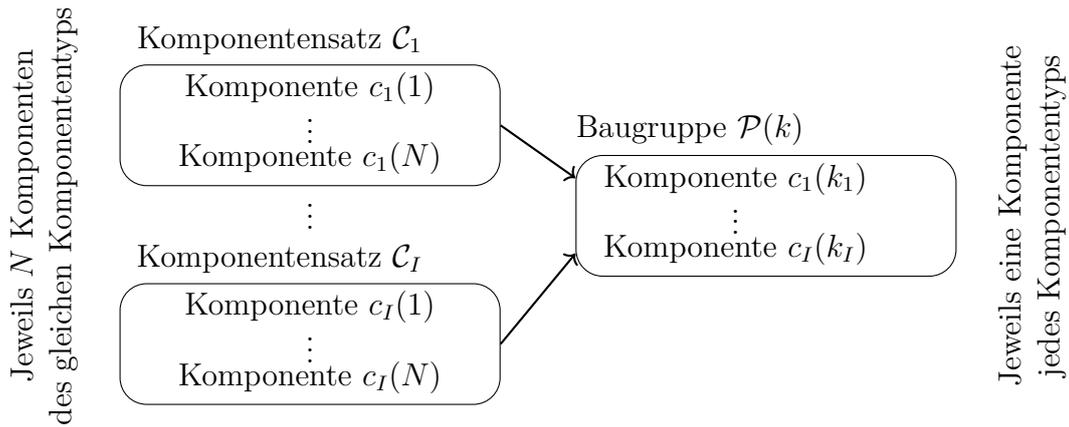


Bild 4.3: Erstellung einer Baugruppe aus gegebenen Komponentensätzen.

Menge  $\mathcal{C}_i = \{c_i(1), c_i(2), \dots, c_i(N)\}$  beschrieben, wobei  $c_i(k)$  mit  $i = 1, 2, \dots, I$  und  $k = 1, 2, \dots, N$  eine Einzelkomponente bezeichnet. Eine Baugruppe  $\mathcal{P}(k) = \{c_1(k_1), c_2(k_2), \dots, c_I(k_I)\}$  mit  $k_i \in \{k \in \mathbb{N} \mid 1 \leq k \leq N\} \forall i = 1, 2, \dots, I$  ist eine spezifische Kombination von Einzelkomponenten und besteht aus je einer Komponente jeden Typs.

Die Aufgabe der Versuchsplanung ist die Erstellung von Kombinationen aller Einzelkomponenten, sodass alle Baugruppen gemeinsam eine optimale Datenverteilung bilden. Bild 4.4 stellt beispielhaft diese Versuchsplanungsaufgabe zur Erstellung von  $N = 3$  Baugruppen aus  $I = 2$  Komponententypen dar. Komponenten werden aus den gegebenen Mengen  $\mathcal{C}_1$  und  $\mathcal{C}_2$  ohne zurücklegen gezogen und zu den Baugrup-

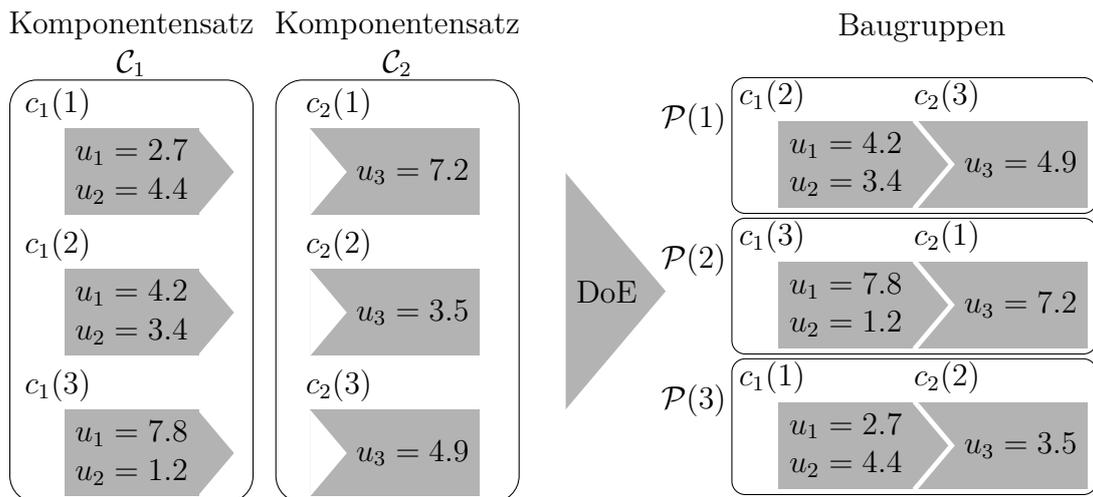


Bild 4.4: Darstellung der Versuchsplanungsaufgabe (DoE) für Baugruppen aus zwei verschiedenen Komponenten mit zufälligen Merkmalsausprägungen und zufälliger Zuordnung der Komponenten zu den Baugruppen [115].

pen  $\mathcal{P}(1)$ ,  $\mathcal{P}(2)$  und  $\mathcal{P}(3)$  kombiniert. Die Komponenten des Typs  $\mathcal{C}_1$  werden hierbei durch zwei Merkmale  $u_1$  und  $u_2$  und die Komponenten in  $\mathcal{C}_2$  durch ein Merkmal  $u_3$  beschrieben. Alle Merkmale der verwendeten Komponenten  $u_1, u_2, u_3$  sind zugleich Merkmale der resultierenden Baugruppe. Die Anzahl an betrachteten Merkmalen eines Komponententyps  $i$  wird im Folgenden mit  $m_i \in \mathbb{N}$  bezeichnet. Aus deren Summe ergibt sich die Dimensionalität des Eingangsraums zu

$$p = \sum_{i=1}^I m_i. \quad (4.1)$$

Jede Baugruppe  $\mathcal{P}(k)$ , die sich aus der Kombination ihrer Einzelkomponenten ergibt, entspricht einem Punkt  $\underline{u}(k) = [u_1(k) \ u_2(k) \ \cdots \ u_p(k)]^T$  mit  $k = 1, 2, \dots, N$  im  $p$ -dimensionalen Eingangsraum, wobei  $u_j(k)$  mit  $j = 1, 2, \dots, p$  die einzelnen Merkmale bezeichnet. Die  $N$  Datenpunkte aller Baugruppen können zeilenweise in einer Datenmatrix

$$\underline{X} = \begin{bmatrix} u_1(1) & u_2(1) & \cdots & u_p(1) \\ u_1(2) & u_2(2) & \cdots & u_p(2) \\ \vdots & \vdots & & \vdots \\ u_1(N) & u_2(N) & \cdots & u_p(N) \end{bmatrix} \quad (4.2)$$

← Datenpunkt  $\underline{u}(2)$

← Merkmal  $u_2$

zusammengefasst werden, wobei exemplarisch eine Spalte und eine Zeile dieser Matrix hervorgehoben sind. Jede Spalte enthält alle Ausprägungen eines Merkmals  $u_j$ , die durch die zur Verfügung stehenden Einzelkomponenten eines Typs festgelegt sind. Mit der Datenmatrix wird das spezielle Versuchsplanungsproblem für kombinierbare Einheiten in eine Standardform überführt. Auf diese Weise wird die Verwendung üblicher Methoden, z. B. zur Bewertung der Datenverteilungen ermöglicht.

### Randbedingungen

Im Gegensatz zu klassischen Versuchsplanungsverfahren ergeben sich bei der Betrachtung von zusammengesetzten Einheiten mehrere Randbedingungen, die berücksichtigt werden müssen, um einen gültigen Versuchsplan zu erhalten. So dürfen in den Versuchspunkten ausschließlich die vorgegebenen Merkmalsausprägungen der Komponenten verwendet werden. Da jede Komponente nur in genau einer Baugruppe verbaut werden kann, definiert diese Merkmalsausprägungen in genau einer

Zeile der Datenmatrix. Bild 4.5 stellt die hieraus resultierenden Einschränkungen anhand eines Montageprozesses von zwei Komponenten, die durch jeweils ein Merkmal beschrieben werden, dar. Ausgehend von einem initialen Versuchsplan in (a) sind alle Permutationen der Komponenten in (b) dargestellt. Hierbei handelt es sich um alle Versuchspunkte, die prinzipiell aus den gegebenen Sätzen an Komponenten erzeugt werden können. Im Beispiel sind keine Merkmalsausprägungen im Bereich um  $u_1 = 0.5$  gegeben, sodass hier in jedem Versuchsplan eine Lücke verbleibt. Ein gültiger Versuchsplan in (c) ist eine Untermenge mit  $N$  Elementen dieser möglichen Menge an Versuchspunkten, wobei jede gegebene Komponente als Randbedingung in nur genau einer Baugruppe verwendet werden kann. Zusätzlich muss durch den Versuchsplanungsalgorithmus sichergestellt werden, dass bei Komponenten  $i$  mit mehreren Merkmalen  $m_i > 1$  die gegebenen Kombinationen an Merkmalsausprägungen erhalten bleiben. Dies ist beispielsweise in Bild 4.4 für den Komponententyp  $\mathcal{C}_1$  mit den Merkmalen  $u_1$  und  $u_2$  notwendig.

## Optimierung

Für die Optimierung von Versuchsplänen können zahlreiche Algorithmen verwendet werden. So wurden beispielsweise D-optimale Versuchspläne mittels lokaler Optimierungsverfahren [81], genetischer Algorithmen [76] oder Simulated Annealing [43]

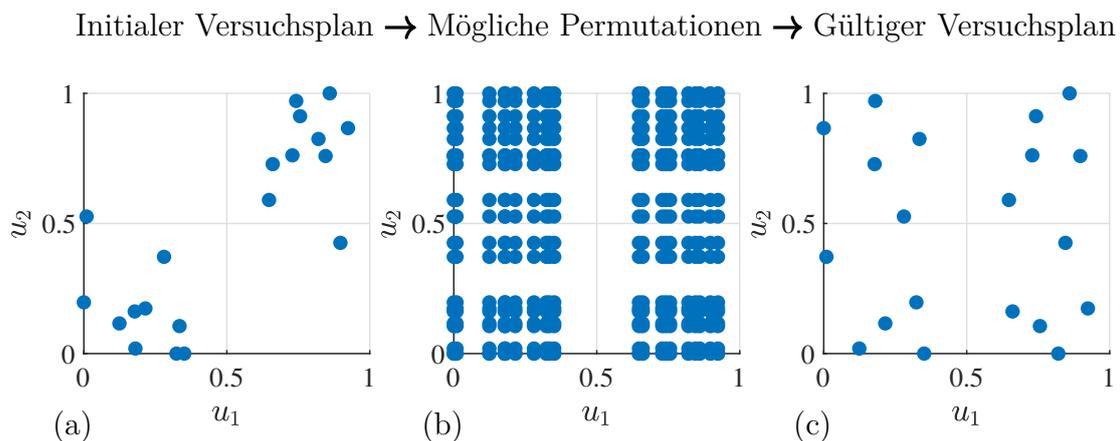


Bild 4.5: Aus einem initialen Versuchsplan (a) mit zwei Komponenten, die jeweils durch ein Merkmal beschrieben werden, werden alle möglichen Permutationen (b) erzeugt. Diese Punkte können prinzipiell auf Basis des gegebenen Satzes an Komponenten realisiert werden. Ein gültiger Versuchsplan (c) ist eine Auswahl dieser Punkte, wobei zu berücksichtigen ist, dass jede Komponente nur in genau einer Baugruppe verbaut werden kann.

erstellt. Diese Verfahren können prinzipiell auch auf das Versuchsplanungsproblem für zusammengesetzte Einheiten angewandt werden. Nachfolgend werden ein lokales (Abschnitt 4.2.1) und ein globales Optimierungsverfahren (Abschnitt 4.2.2) zur Erstellung von Versuchsplänen für zusammengesetzte Einheiten beschrieben. Beide Verfahren zielen darauf ab, raumfüllende Versuchspläne zu erstellen, wobei die zuvor genannten Randbedingungen berücksichtigt werden.

### 4.2.1 Lokaler Suchalgorithmus

Lokale Suchalgorithmen stellen eine häufig verwendete Klasse von Optimierungsverfahren für kombinatorische Optimierungsprobleme dar. Diese heuristischen Verfahren zielen darauf ab, ein hinreichend gutes lokales Optimum in akzeptabler Rechenzeit zu finden. Hierzu werden ausgehend von einer Startlösung benachbarte Lösungen systematisch durchsucht und bei einer Verbesserung des Optimierungskriteriums übernommen. Dieses Vorgehen wird iterativ fortgesetzt, bis beispielsweise keine signifikante Verbesserung des Optimierungskriteriums mehr erzielt wird [75]. Eine Möglichkeit, um benachbarte Lösungen zu erzeugen, ist die Durchführung von Tauschvorgängen. So tauscht der in Abschnitt 2.4.2 beschriebene EDLS-Algorithmus Koordinaten von Versuchspunkten, um einen gegebenen Versuchsplan zu optimieren. Da die Merkmalsausprägungen bei diesen Koordinatentauschvorgängen erhalten bleiben, ist dieser Algorithmus direkt zur Erstellung von Versuchsplänen für zusammengesetzte Einheiten geeignet. Allerdings darf dann jede Komponente nur durch exakt ein Merkmal  $m_i = 1 \forall i = 1, 2, \dots, I$  beschrieben werden.

#### Erweiterung auf mehrere Merkmale pro Komponente

Für viele reale Anwendungen ist die Beschränkung auf ein Merkmal pro Komponente unzureichend. Aus diesem Grund wird der Koordinatentauschmechanismus im EDLS-Algorithmus zu einem Komponententauschmechanismus erweitert. Bei einem Tauschvorgang wird für alle  $m_i$  Merkmale einer Komponente dieselbe Tauschopeation durchgeführt, sodass die durch eine Komponente definierte Kombination an Merkmalsausprägungen erhalten bleibt. In Bild 4.6 ist eine Abfolge von Komponententauschvorgängen und deren Auswirkung auf die Datenmatrix ( $\underline{X}_1$ ,  $\underline{X}_2$  und  $\underline{X}_3$ ) dargestellt. Zusätzlich ist die minimale Punktdistanz  $d_{\min}$  (siehe Gleichung 2.21) angegeben, die durch den EDLS-Algorithmus maximiert wird. Das Beispiel basiert

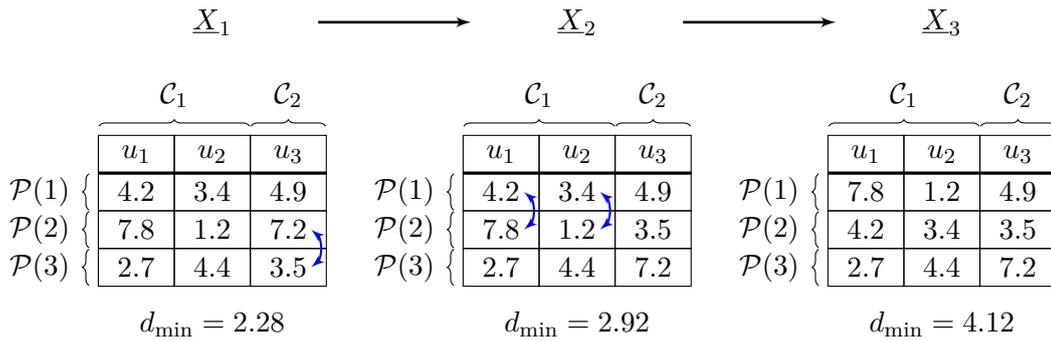


Bild 4.6: Beispielhafte Tauschvorgänge von Komponenten und deren Auswirkungen auf die Datenmatrix für den in Bild 4.4 dargestellten Montageprozess.

auf dem in Bild 4.4 beschriebenen Montageprozess und dem dort dargestellten Versuchsplan. Im ersten Schritt erfolgt ein Tausch einer Komponente des Typs  $\mathcal{C}_2$  zwischen den Baugruppen  $\mathcal{P}(2)$  und  $\mathcal{P}(3)$ . Da hiervon mit  $u_3$  nur ein Merkmal betroffen ist, entspricht der Komponententausch einem Koordinatentausch des ursprünglichen EDLS-Algorithmus. Im zweiten Schritt wird eine Komponente des Typs  $\mathcal{C}_1$  zwischen den Baugruppen  $\mathcal{P}(1)$  und  $\mathcal{P}(2)$  getauscht, die durch  $m_1 = 2$  Merkmale beschrieben wird. Bei diesem Tauschvorgang werden die Ausprägungen der Merkmale  $u_1$  und  $u_2$  gemeinsam getauscht. Beide Optimierungsschritte führen zu einer Vergrößerung der minimalen Punktdistanz  $d_{\min}$  und bewirken somit eine Verbesserung des Versuchsplans.

### Demonstration an einem einfachen Montageprozess

Eine Anwendung des lokalen Suchalgorithmus für zusammengesetzte Einheiten ist am Beispiel eines Montageprozesses mit  $I = 2$  Komponententypen, die durch jeweils zwei Merkmale  $m_1 = m_2 = 2$  beschrieben werden, in Bild 4.7 dargestellt. Alle  $N = 8$  Versuchspunkte sind in (a) entlang einer Diagonalen im Eingangsraum initialisiert, sodass dieser Versuchsplan in Bezug auf eine gleichmäßige Abdeckung des Eingangsraums einen ungünstigen Fall darstellt. In (b) ist der resultierende Versuchsplan nach der Optimierung mittels des lokalen Suchalgorithmus zu sehen. Die grau hinterlegten Streudiagramme können bei der Optimierung nicht verändert werden, da diese ausschließlich Merkmale eines Komponententyps darstellen und jede Einzelkomponente eine feste Kombination dieser Merkmalsausprägungen definiert. In allen anderen Streudiagrammen ist die gesteigerte Gleichförmigkeit der Datenverteilung durch die Anwendung des lokalen Suchalgorithmus sichtbar. Je mehr Merkmale einer Komponente in die Versuchsplanung miteinbezogen werden, desto mehr feste Kom-

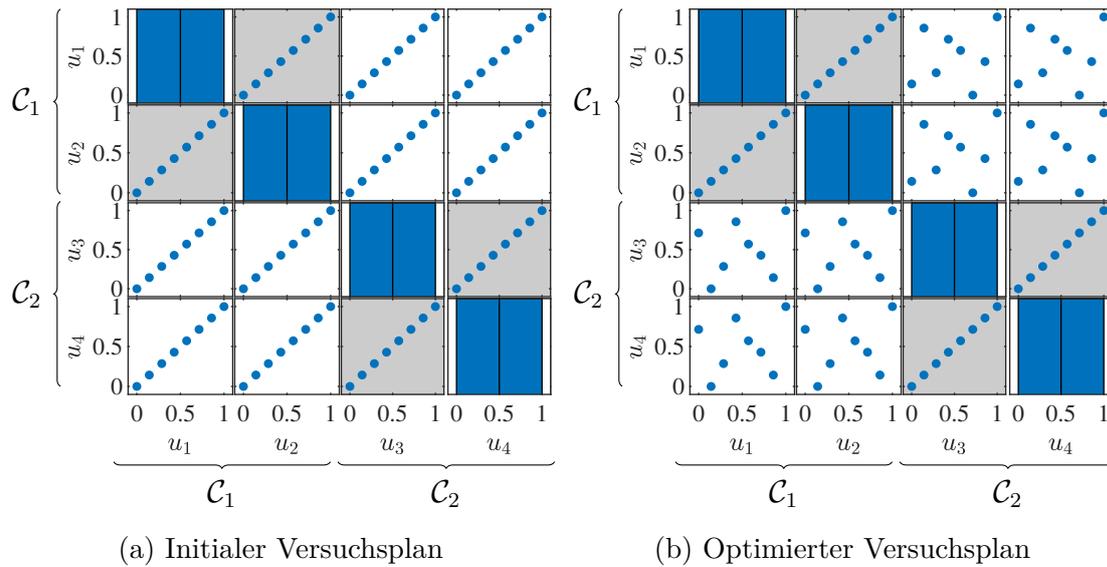


Bild 4.7: Matrixplots zum Vergleich (a) eines ungünstig initialisierten Versuchsplans für einen Montageprozess und (b) des Versuchsplans nach der Optimierung [116]. Es werden zwei Komponententypen  $\mathcal{C}_1$  und  $\mathcal{C}_2$  mit jeweils zwei Merkmalen betrachtet. Die grau hinterlegten Felder enthalten nur Merkmale eines Komponententyps und können daher nicht verändert werden.

binationen zwischen Merkmalen (grau hinterlegte Felder) sind vorhanden, die nicht durch die Versuchsplanung beeinflusst werden können.

### 4.2.2 Globaler Suchalgorithmus mittels genetischer Optimierung

In diesem Abschnitt wird ein *genetischer Algorithmus* (GA) vorgestellt, der zur Klasse der globalen Optimierungsverfahren zählt. Eine grundlegende Version dieses Algorithmus wurde im Rahmen der Masterarbeit [44] implementiert. Das Vorgehen orientiert sich an [4], in dem ein GA zur Erstellung D-optimaler Versuchspläne für Montageprozesse verwendet wird. Eine Beschreibung der Grundlagen der genetischen Optimierung ist in Anhang A.3 zu finden, während in diesem Abschnitt nur die Anpassungen an das betrachtete Versuchsplanungsproblem beschrieben werden.

Bei dieser Anwendung der Versuchsplanung für zusammengesetzte Einheiten beschreibt jedes der  $A$  Individuen  $S_l$  mit  $l = 1, 2, \dots, A$  einen vollständigen Versuchsplan und gibt somit vor, welche Kombinationen von Einzelkomponenten zu Baugruppen zusammengefügt werden. Es existieren verschiedene Möglichkeiten, dieses Optimierungsproblem für den GA zu beschreiben. Da es sich bei der Zusammenstellung

der Komponenten um ein kombinatorisches Problem handelt, wird eine Beschreibung als Permutationen gewählt [64]. Hierbei gibt eine Sequenz natürlicher Zahlen die in den Baugruppen verwendeten Einzelkomponenten an. Bild 4.8 stellt den Versuchsplan aus Bild 4.4 mit der für den genetischen Algorithmus gewählten Codierung dar. Hierzu werden in (a) alle Komponenten in den gegebenen Komponentensätzen durchnummeriert. Durch Anwendung der Permutationen (b) ergibt sich eine geänderte Reihenfolge der Komponenten, die in (c) der Reihe nach den  $N$  Baugruppen zugeordnet werden. Die Permutationen dienen als Gensequenz im GA und werden in Form eines  $N$ -Tupels  $s_i$  für jeden Komponententyp  $i = 1, 2, \dots, I$  beschrieben. Ein Individuum ist über  $I$  solcher Sequenzen definiert, die wiederum zu einem Tupel  $S_I = (s_1, s_2, \dots, s_I)$  zusammengefasst werden. Das Individuum in Bild 4.8 ergibt sich somit zu  $S = ((2, 3, 1), (3, 1, 2))$ .

### Selektion von Individuen

Zur Auswertung der Fitnessfunktion  $F(S_I)$  eines Individuums wird zunächst die zum Individuum zugehörige Datenmatrix  $\underline{X}_I$  generiert. Diese ist durch die Permutationen der gegebenen Komponenten mit den zugehörigen Merkmalsausprägungen eindeutig beschrieben. Auf Grundlage dieser Datenmatrix wird anschließend die minimale euklidische Distanz  $d_{\min}$  im Versuchsplan bestimmt. Der GA maximiert diese Distanz, sodass dasselbe Optimierungskriterium wie bei dem lokalen Suchalgorithmus vorliegt. Wird durch eine genetische Operation ein Individuum erzeugt, das einen verbesserten Versuchsplan beschreibt, führt dies typischerweise nur zu kleinen Wertänderungen von  $d_{\min}$ . Somit wäre bei Verwendung der üblichen *Roulette Wheel Selection* (siehe Anhang A.3) nur eine geringfügig erhöhte Selektionswahrscheinlichkeit dieses Individuums gegenüber den weiteren Individuen in einer Generation gegeben.

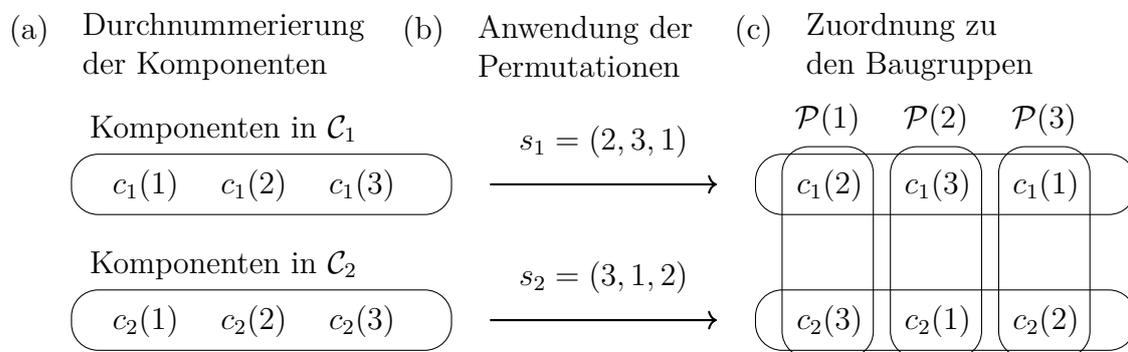


Bild 4.8: Codierung des Versuchsplanungsproblems für die genetische Optimierung am Beispiel von  $N = 3$  Baugruppen aus  $I = 2$  Komponententypen.

Die erzielte Verbesserung würde daher nicht in dem benötigten Maße fortbestehen. Um das Problem des geringen Wertebereichs der Fitnessfunktion zu beheben, wird die *Tournament Selection* genutzt, in der die Fitnesswerte mehrerer Individuen verglichen werden. Um ein Individuum zu selektieren, werden zunächst  $q$  Individuen zufällig aus der aktuellen Generation ausgewählt. Unter diesen Individuen wird anschließend dasjenige mit dem höchsten Fitnesswert bestimmt und einer genetischen Operation unterzogen [37].

### Genetische Operationen

Die genetische Operation der Mutation ermöglicht durch zufällige Veränderung von Individuen ein Verlassen lokaler Optima. Typischerweise werden nur geringfügige Veränderungen vorgenommen, um einen Großteil der Informationen zu bewahren, die in den vorherigen Iterationsschritten in das Individuum eingeflossen sind. Hierzu wird eine vertauschende Mutation verwendet, bei der zwei zufällige Zahlen innerhalb eines zufällig ausgewählten Tupels  $s_i$ , das eine Permutation der Einzelkomponenten eines Typs beschreibt, miteinander vertauscht werden. Durch das Vertauschen wird sichergestellt, dass stets eine gültige Permutation vorliegt [118].

Bei der Rekombination werden zwei Individuen  $S_A$  und  $S_B$  einer Generation kombiniert, wobei jedes Individuum Teile seiner Gensequenz weitergibt. Die hier verwendete Rekombination bezieht sich auf einen zufällig ausgewählten Komponententyp, sodass sich neue Permutationen  $s_i$  für die betrachteten Individuen ergeben. Bei dem vorliegenden kombinatorischen Problem muss sichergestellt werden, dass jede resultierende Gensequenz eine gültige Permutation darstellt und somit jede Komponente genau einer Baugruppe zugewiesen wird. Zu diesem Zweck wird das *Partially-Matched-Crossover-Verfahren* (PMX) [59] verwendet.

Exemplarisch ist eine Rekombination mittels PMX für die Gensequenz  $s_1$  des Komponententyps  $i = 1$  in Bild 4.9 dargestellt. Im ersten Schritt (a) wird hierzu ein zufälliger Bereich der Gensequenz (grau) gewählt. Hieraus ergibt sich eine Zuordnungsvorschrift (unten), deren Anwendung in (b) zu einem Vertauschen der Elemente in diesem Bereich führt. Die resultierenden Tupel weisen Duplikate auf und stellen somit keine Permutationen dar. Um hieraus zulässige Permutationen zu erzeugen, wird in (c) die zuvor erstellte Zuordnung verwendet, um die mehrfach auftretenden Elemente außerhalb des Bereichs zu ersetzen. Im Beispiel wird so das erste Element

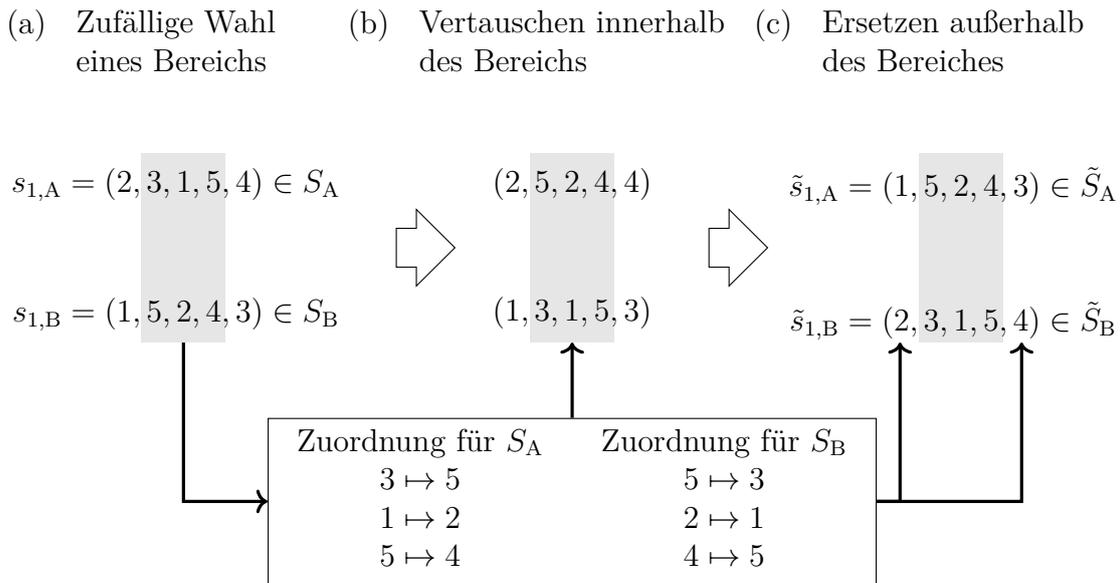


Bild 4.9: Beispiel (a) für die Rekombination zweier Individuen  $S_A$  und  $S_B$  mittels PMX, wobei der zufällig gewählte Bereich die Elemente 2-4 umfasst. Nach (b) dem Vertauschen der Elemente in diesem Bereich werden in (c) Elemente außerhalb des Bereiches ersetzt, um zulässige Permutationen zu erzeugen.

der Sequenz  $s_{1,A}$  gemäß der Zuordnung  $2 \mapsto 1$  ausgetauscht. Elemente an einer Position können auch mehrmals hintereinander ersetzt werden, um Duplikate aufzulösen. So werden auf das fünfte Element von  $s_{1,A}$  die Zuordnungen  $4 \mapsto 5$  und  $5 \mapsto 3$  nacheinander angewandt, sodass keine Duplikate mehr vorliegen. Die resultierenden Individuen  $\tilde{S}_A$  und  $\tilde{S}_B$  werden Teil der Nachfolgeneration.

Die Wahl der Hyperparameter hat bei einem GA einen großen Einfluss auf das Konvergenzverhalten und die erzielte Güte. Die wichtigsten dieser Hyperparameter sind in Tabelle 4.1 zusammengefasst, wobei die gewählten Werten der Literatur entnommen wurden.

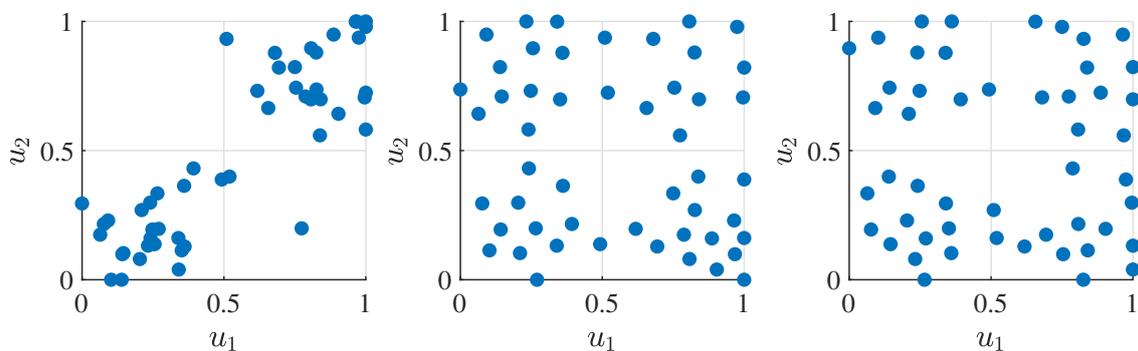
Tabelle 4.1: Verwendete Hyperparameter des genetischen Algorithmus

Hyperparameter	Wert	Quelle
Anzahl der Individuen	$A = 50$	[4]
Mutationswahrscheinlichkeit	$p_M = 0.02$	[59]
Rekombinationswahrscheinlichkeit	$p_K = 0.6$	[59]
Reproduktionswahrscheinlichkeit	$p_P = 0.38$	[59]
Anzahl der Iterationen	$W_G = 2000$	[4]
Betrachtete Individuen in der Tournament Selection	$q = 3$	[66]

### 4.2.3 Vergleich der Optimierungsalgorithmen

Die vorgestellten Optimierungsalgorithmen werden anhand einer initialen Datenverteilung der Komponente verglichen, in der die Versuchspunkte konzentriert in zwei Clustern vorliegen. Diese Datenverteilung stellt eine realitätsnahe Annahme für das Verhalten industrieller Prozesse zur Herstellung der Einzelkomponenten dar. So weisen Einzelkomponenten, die unter ähnlichen Bedingungen gefertigt werden, häufig eine geringe Streuung auf und bilden somit ein Cluster. Mehrere solcher Cluster entstehen durch Einflüsse von zum Beispiel Werkzeugwechseln oder geänderten Chargen des Rohmaterials. Die angenommene Datenverteilung besteht aus zwei Clustern mit jeweils  $N/2$  Datenpunkten, die durch multivariate Normalverteilungen  $\mathcal{N}(\underline{\mu}_j, \sigma_C^2 \underline{I})$  mit  $j = 1, 2$  und  $\sigma_C = 0.15$  beschrieben werden. Es werden zwei unterschiedliche Erwartungswertvektoren  $\underline{\mu}_1 = [0.2 \ 0.2]^T$  und  $\underline{\mu}_2 = [0.8 \ 0.8]^T$  verwendet, sodass sich Cluster in gegenüberliegenden Ecken des Eingangsraums ergeben.

Die beschriebene Datenverteilung mit der Dimensionalität von  $p = 2$  wird zum Vergleich der Algorithmen verwendet, wobei zwei Komponententypen mit jeweils einem Merkmal betrachtet werden. Bild 4.10a zeigt exemplarisch eine initiale Datenverteilung für  $N = 50$  Versuchspunkte. Die resultierenden optimierten Datenverteilungen sind in Bild 4.10b für den lokalen Suchalgorithmus und in Bild 4.10c für den GA dargestellt. Bei allen Datenverteilungen ist im Vergleich zu den Ecken eine geringere Punktdichte in der Mitte des Eingangsraums sichtbar, die aus den gegebenen Merkmalsausprägungen der Komponenten resultiert. Werden die mittels Optimierung erzeugten Datenverteilungen verglichen, ist tendenziell eine höhere Gleichförmigkeit der in (b) durch den lokalen Suchalgorithmus erzeugten Datenverteilung erkennbar.

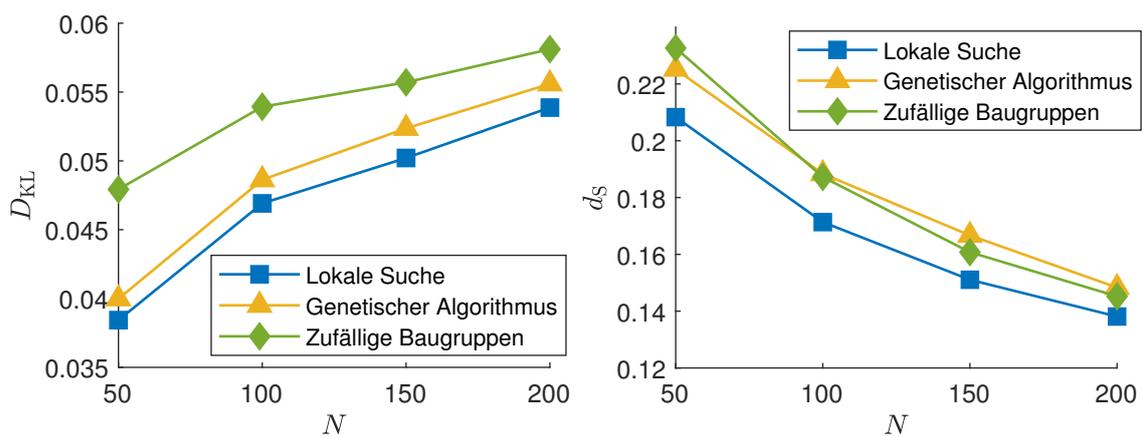


(a) Initiale Datenverteilung (b) Lokaler Suchalgorithmus (c) Genetischer Algorithmus

Bild 4.10: Vergleich der Datenverteilungen vor und nach der Optimierung. Aus (a) der initialen Datenverteilung werden (b) und (c) erzeugt.

Um dieses Ergebnis zu quantifizieren und die Optimierungsverfahren hinsichtlich der erzielten Raumabdeckung zu vergleichen, werden die in Abschnitt 2.3 beschriebenen Bewertungsmaße verwendet. Bild 4.11 zeigt die mit beiden Algorithmen erzielten Ergebnisse für verschiedene Umfänge der Versuchspläne, wobei mit  $D_{\text{KL}}$  und  $d_S$  die vom Optimierungskriterium unabhängigen Bewertungsmaße dargestellt sind. Als Vergleich wird zusätzlich eine zufällige Zusammenstellung der Komponenten zu Baugruppen betrachtet. Eine solche zufällige Zusammenstellung ist ein übliches Vorgehen für reale Versuche, wie sie z. B. zur Überprüfung von Toleranzsimulationen durchgeführt werden, und stellt damit eine Referenzmethode aus der industriellen Praxis dar [8].

Mit steigender Anzahl  $N$  an Datenpunkten ist für alle Verfahren eine Zunahme von  $D_{\text{KL}}$  sichtbar. Dies ist auf die verwendete initiale Datenverteilung zurückzuführen, in der Cluster von Datenpunkten vorliegen. Die Datenpunkte in diesen Clustern werden durch die Optimierung hauptsächlich über die Randbereiche des Eingangsraums verteilt. Je mehr Punkte in den initialen Clustern liegen, desto größer wird die Punktdichte in diesen Randbereichen, während die Punktdichte in der Mitte des Eingangsraums klein bleibt. Für  $D_{\text{KL}}$  sind bei beiden Optimierungsverfahren kleinere Werte und damit eine höhere Gleichförmigkeit der Datenverteilung als bei einer zufälligen Zusammenstellung der Baugruppen sichtbar, wobei die lokale Suche die besten Ergebnisse erzielt. Das Kriterium  $d_S$  beschreibt die größte Lücke im Eingangsraum, die sich mit steigender Anzahl  $N$  an Datenpunkten verringert. Auch in Bezug



(a) Beurteilung der Gleichförmigkeit mittels  $D_{\text{KL}}$ . (b) Beurteilung von Lücken im Eingangsraum mittels  $d_S$ .

Bild 4.11: Vergleich der lokalen Suche und des GA bei verschiedenen Anzahlen an Datenpunkten. Mittelwerte über 100 Wiederholungen sind dargestellt.

auf dieses Kriterium erzielt der lokale Suchalgorithmus die geringsten Werte und damit die besten Ergebnisse.

Zusammenfassend sind beide Algorithmen zur Erstellung von Versuchsplänen für zusammengesetzte Einheiten geeignet und führen zu besseren Datenverteilungen als zufällige Versuchspläne. Hierbei erzielt der lokale Suchalgorithmus insgesamt bessere Ergebnisse als der GA, was bereits in [97] bei der Erstellung von D-optimalen Versuchsplänen für Montageprozesse beobachtet wurde. Dies kann durch die unmittelbare Sicherstellung gültiger Versuchspläne durch die Tauschoperationen bei der lokalen Suche erklärt werden. Beim GA erfolgt hingegen eine nachträgliche Korrektur der Permutationen durch die PMX-Rekombination, was die Zufälligkeit steigert. Das Verhalten des GA kann prinzipiell durch geeignete Hyperparameter weiter verbessert werden, jedoch stellt dieser Anpassungsbedarf einen entscheidenden Nachteil in der praktischen Anwendung dieser Versuchsplanungsmethode dar. Nachfolgend wird aufgrund der besseren Ergebnisse und der einfacheren Anwendung die lokale Suche verwendet.

### 4.3 Verbesserte Schätzung von Rauscheinflüssen

Die raumfüllenden Versuchspläne, die mittels der in diesem Kapitel beschriebenen Methodik erstellt werden, dienen zur Schätzung von datenbasierten Modellen. Hierbei ist es das Ziel, die funktionalen Zusammenhänge zwischen den Merkmalen der Komponenten und den Eigenschaften der Baugruppen im gesamten Eingangsraum abzubilden. Die Auswirkungen verschiedener Versuchspläne auf die Modellgüte werden nachfolgend anhand von Gaußprozessmodellen (GPM) untersucht. Werden deterministische Prozesse modelliert, wie es beispielsweise bei vielen Computersimulationen der Fall ist, sind raumfüllende Versuchspläne eine typische Wahl [60]. Handelt es sich bei dem zu modellierenden Prozess jedoch um einen stochastischen Prozess, können ausschließlich raumfüllende Versuchspläne zu einer fehlerhaften Schätzung der Rauschvarianz  $\hat{\sigma}_n^2$  führen [61].

Eine häufig genutzte Möglichkeit, die Modellqualität beim Vorliegen stochastischer Einflüsse zu steigern, ist die wiederholte Vermessung von Punkten (*Replikationen*). Beispielsweise wird in [3] ein Verfahren für GPM beschrieben, um eine festgelegte Anzahl an Wiederholungen auf gegebene Versuchspunkte zu verteilen. Bei Versuchen mit zusammengesetzten Einheiten sind exakte Wiederholungen von Versuchspunkten durch die individuellen Merkmalsausprägungen der Komponenten im Allgemeinen

nicht möglich. Stattdessen ist es das Ziel des nachfolgend beschriebenen Vorgehens, eine vorgegebene Anzahl an möglichst ähnlichen Punkten zu erzeugen. Wie in [117] beschrieben, hat ein Cluster ähnlicher Punkte einen vergleichbaren Effekt auf den Vorhersagefehler eines GPM wie eine exakte Wiederholung von Versuchspunkten.

### Erzeugung von näherungsweise Wiederholpunkten

Zur Erzeugung möglichst ähnlicher Versuchspunkte wird anstatt des bisher verwendeten Maximin-Kriteriums (siehe Gleichung (2.28)) mit

$$\min_{\mathcal{D}}(d_{\min}) = \min_{\mathcal{D}} \left( \min_{\underline{u}(k), \underline{u}(l) \in \mathcal{D}, k \neq l} (d(\underline{u}(k), \underline{u}(l))) \right) \quad (4.3)$$

ein Minimin-Kriterium verwendet. Bild 4.12 stellt in (a) einen initialen Versuchsplan für zwei Komponententypen dar. Dieser Versuchsplan wird mit dem lokalen Suchalgorithmus unter Verwendung des Minimin-Kriteriums optimiert und führt zu dem Versuchsplan in (b), der Cluster von Versuchspunkten aufweist. Aus diesen Clustern sollen die  $K$  Punkte ausgewählt werden, die sich am ähnlichsten sind. Hierzu wird für jeden Versuchspunkt die Summe der euklidischen Abstände zu dessen  $K - 1$  nächsten Nachbarn bestimmt. Diejenigen Punkte mit der geringsten Summe dieser Abstände werden als die  $K$  ähnlichsten Punkte im Versuchsplan betrachtet (grüne Markierung). Die ähnlichsten Punkte bleiben während der in (c) folgenden Optimierung

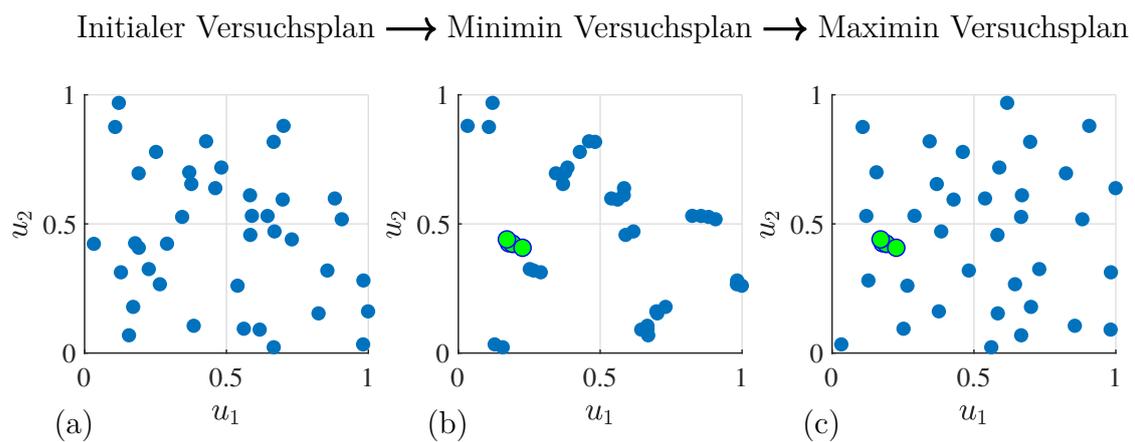


Bild 4.12: Vorgehen zur Erzeugung eines raumfüllenden Versuchsplans mit ähnlichen Versuchspunkten. Der initiale Versuchsplan (a) mit  $N = 40$  Punkten wird in (b) mittels Minimin-Kriterium optimiert.  $K = 5$  Punkte (grün) bleiben während der anschließenden Maximin-Optimierung in (c) unverändert.

mit dem Maximin-Kriterium unverändert. Auf diese Weise entsteht ein raumfüllender Versuchsplan mit  $K$  näherungsweise wiederholten Versuchspunkten.

### Demonstration der näherungsweisen Wiederholpunkte an einem Testprozess

Um den Nutzen dieser ähnlichen Punkte für die Modellgüte zu untersuchen, wird ein Testprozess definiert. Hierzu wird die multivariate Gaußfunktion

$$f(\underline{u}) = \exp\left(-\frac{1}{2} \sum_{j=1}^p \frac{(u_j - 0.5)^2}{0.3^2}\right) + n \quad (4.4)$$

als Testfunktion verwendet. Zur Einbeziehung stochastischer Einflüsse wird ein additives weißes Rauschen  $n \sim \mathcal{N}(0, \sigma_n)$  berücksichtigt. Es werden  $I = 3$  Komponenten mit jeweils einem Merkmal  $m_1 = m_2 = m_3 = 1$  und  $N = 90$  Baugruppen betrachtet. Die verschiedenen Trainingsdatensätze werden zur Erstellung von GPM genutzt.

Die aus den unterschiedlichen Versuchsplänen resultierenden Modelle sind in Tabelle 4.2 für verschiedene Standardabweichungen des Rauschens gegenübergestellt. Sie werden anhand des RMSE auf Basis einer Sobol-Sequenz mit  $N_{\text{test}} = 2^{14}$  Testdatenpunkten und der geschätzten Standardabweichung des Rauschens  $\hat{\sigma}_n$  verglichen. Diese Standardabweichung des Rauschens wird bei der Modellerstellung zusammen

Tabelle 4.2: Vergleich der Modellgüte und der geschätzten Standardabweichung des Rauschens von GPM, die auf Basis verschiedener Versuchsplänen mit und ohne Erzeugung ähnlicher Versuchspunkte erstellt wurden. Alle Ergebnisse sind über 100 Wiederholungen gemittelt.

Rauschen $\sigma_n$	Erzeugte Anzahl ähnlicher Punkte	RMSE	Geschätztes Rauschen $\hat{\sigma}_n$
0.01	keine	0.0176	0.0040
	6	0.0183	0.0081
	3 · 2	0.0177	0.0075
0.025	keine	0.0358	0.0174
	6	0.0359	0.0200
	3 · 2	0.0355	0.0198
0.05	keine	0.0637	0.0423
	6	0.0639	0.0432
	3 · 2	0.0637	0.0425

mit weiteren Hyperparametern wie den Standardabweichungen der Kernelfunktionen aus den Trainingsdaten geschätzt, indem die Log-Marginal-Likelihood maximiert wird (siehe Abschnitt 2.2.3). Die verwendeten Versuchspläne unterscheiden sich in Bezug auf die erzeugten ähnlichen Punkte, wobei entweder keine oder  $K = 6$  solcher Punkte betrachtet werden. Zusätzlich ist eine Aufteilung der ähnlichen Punkte auf drei Cluster mit jeweils 2 Punkten ( $3 \cdot 2$ ) gegenübergestellt. Obwohl die ähnlichen Punkte zu einer ungleichmäßigeren Abdeckung des Eingangsraums führen, weisen alle Modelle vergleichbare Werte des RMSE auf. Bei allen Modellen, die unter Nutzung von Versuchsplänen mit ähnlichen Datenpunkten erstellt wurden, liegt die geschätzte Standardabweichung des Rauschens näher am wahren Wert. Somit führen die ähnlichen Datenpunkte zu robusteren und interpretierbarer Modellen.

Die Verteilung der ähnlichen Datenpunkte auf mehrere Cluster führt insgesamt zu geringfügig besseren Ergebnissen bezogen auf den RMSE, was durch die gleichmäßigere Verteilung der Datenpunkte und die größere Ähnlichkeit zwischen den Punkten in den Clustern zu erklären ist. Bei mehreren kleinen Clustern ist anzunehmen, dass sich die Punkte jeweils ähnlicher sind als bei einem einzelnen großen Cluster. Einem großen Cluster müssen zur Erzielung der vorgegebenen Anzahl an Punkten auch vergleichsweise weit entfernte Punkte zugeordnet werden.

Neben der realistischeren Schätzung des Rauschens bieten die ähnlichen Datenpunkte dieselben Vorteile bei der praktischen Umsetzung der Experimente wie wiederholt vermessene Zentralpunkte in klassischen Versuchsplänen, wie z. B. Central Composite Designs. So kann unter anderem zeitvariantes Prozessverhalten während der Versuche detektiert werden, indem die einzelnen Wiederholungsmessungen zeitlich über die Versuchsreihe verteilt werden. Außerdem kann eine Rauschvarianz aus den Wiederholungsmessungen geschätzt werden [77]. Ist eine größere Anzahl an ähnlichen Datenpunkten in mehreren Clustern über den Eingangsraum verteilt, kann für jedes Cluster eine eigene Rauschvarianz ermittelt und somit zusätzlich die Unabhängigkeit des Rauschens von den Eingangsgrößen überprüft werden.

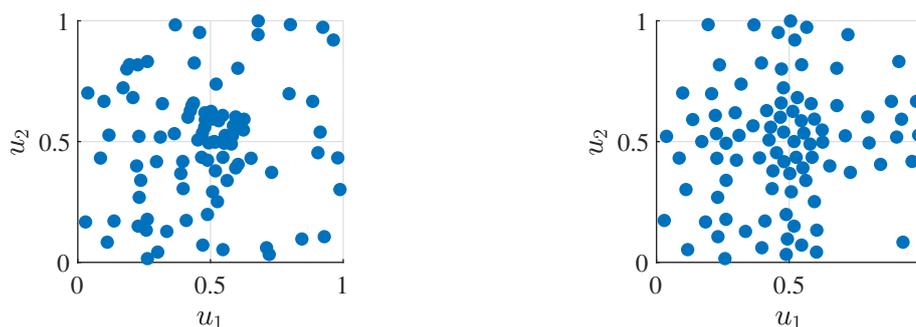
## 4.4 Reduzierte Versuchspläne mittels Punktselektion

Das in diesem Kapitel beschriebene Verfahren ermöglicht es, aus einem gegebenen Satz an Komponenten eine im Sinne der Raumabdeckung bestmögliche Datenverteilung zu erzeugen. Die erzielbare Gleichförmigkeit der Datenverteilung hängt dabei

jedoch stark von den zur Verfügung stehenden Komponenten und deren Merkmalsausprägungen ab. Aus industriellen Prozessen resultieren häufig Komponenten mit einer geringen Streuung der Merkmalsausprägungen, da z. B. bei der Herstellung dasselbe Umformwerkzeug verwendet wird. Ein Versuchsplan für zwei Komponententypen mit jeweils einem Merkmal ist in Bild 4.13a dargestellt, wobei ca. die Hälfte der Komponenten jedes Typs geringe Streuungen aufweisen. Die Datenverteilung resultiert aus der Überlagerung einer Gleichverteilung und einer Normalverteilung und führt zu einer Konzentration von Punkten in der Mitte des Eingangsraums. Durch die Optimierung dieses Versuchsplans werden die Versuchspunkte gleichförmiger verteilt (siehe Bild 4.13b). Trotzdem verbleiben zahlreiche Versuchspunkte in der Mitte des Eingangsraums und weisen somit eine hohe Ähnlichkeit zueinander auf. Eine Vermessung solcher Punkte kann sinnvoll sein, wenn signifikantes Prozessrauschen vorliegt, um eine bessere Schätzung dieses Rauschens zu ermöglichen (siehe Abschnitt 4.3). Andernfalls wird der Versuchsaufwand erhöht, ohne nennenswerte Informationen über die funktionalen Zusammenhänge des betrachteten Prozesses zu erlangen.

### Verfahren zur Punktselektion

Ziel des in diesem Abschnitt beschriebenen Verfahrens zur Punktselektion ist es, die informativsten Punkte aus einem gegebenen Versuchsplan auszuwählen. Bezogen auf einen Montageprozess werden somit Untermengen der gegebenen Komponentensätze in einem Versuch zu Baugruppen verbaut. Ein ähnliches Vorgehen ist in [2]



(a) Initialer Versuchsplan der durch die Überlagerung einer Normal- und einer Gleichverteilung generiert wurde

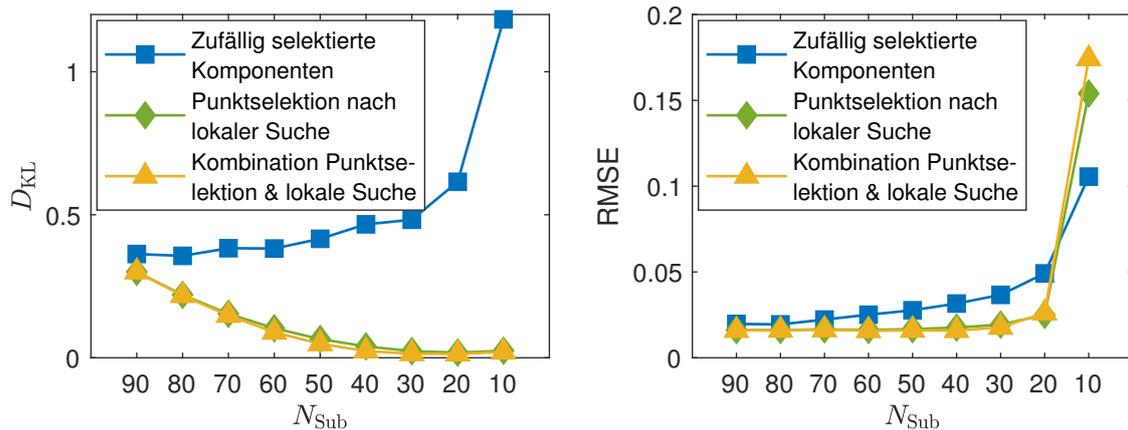
(b) Ergebnis des lokalen Suchalgorithmus zur Optimierung des initialen Versuchsplans

Bild 4.13: Einschränkungen der erzielbaren Gleichförmigkeit durch die gegebenen Komponenten. Die Punkte in der Mitte des Eingangsraums können nur bedingt gleichförmiger verteilt werden.

am Beispiel von Bauteilen beschrieben, die einer zerstörenden Prüfung unterzogen werden sollen. Hierbei wird ein D-optimaler Versuchsplan mittels eines Austauschalgorithmus erzeugt, der Bauteile auf Grundlage ihrer gemessenen Merkmale aus einem großen Kandidatensatz auswählt. Im Gegensatz zum Versuchsplanungsproblem für zusammengesetzte Einheiten sind die Punkte in diesem Kandidatensatz fest vorgegeben und können nicht durch weitere Operationen wie dem Tauschen von Komponenten verändert werden. Um mittels Punktselektion einen raumfüllenden Versuchsplan zu erzeugen, wird das in Abschnitt 2.3.3 beschriebene Punktselektionsverfahren verwendet. Dieses vergleicht hierarchisch die Nächste-Nachbarn-Distanzen der Versuchspunkte und entfernt sequentiell denjenigen Punkt, der zuerst ein Minimum bei einer dieser Distanzen aufweist. Somit kann die Gleichförmigkeit von Versuchsplänen, insbesondere bei ungünstigen Komponentensätzen wie dem in Bild 4.13a gezeigten, gesteigert werden. In der praktischen Anwendung führt dieses Vorgehen zur Verringerung des Versuchsaufwands bei möglichst geringer Reduzierung des Informationsgehalts der Versuchsdaten.

### Demonstration der Punktselektion am Testprozess

Bild 4.14 zeigt am Beispiel der Datenverteilung aus Bild 4.13a und dem Testprozess in Gleichung (4.4) den Nutzen der Punktselektion. Der RMSE wird hierbei auf Basis einer Sobol-Sequenz mit  $N_{\text{test}} = 2^{14}$  Testdatenpunkten ermittelt. Beginnend mit



(a) Vergleich der Gleichförmigkeit der reduzierten Versuchspläne mittels  $D_{\text{KL}}$  (b) Vergleich der aus den reduzierten Versuchsplänen resultierenden Modellfehler

Bild 4.14: Gleichförmigkeit der Versuchspläne und resultierende Modellfehler für unterschiedliche Anzahlen an selektierten Punkten  $N_{\text{Sub}}$ . Alle Ergebnisse sind über 100 Wiederholungen gemittelt.

einem optimierten Versuchsplan für 90 Baugruppen wird fortlaufend die Anzahl der Versuchspunkte mittels Punktselektion reduziert. Zum Vergleich ist ein Versuchsplan mit  $N_{\text{Sub}}$  zufällig zusammengestellten Baugruppen gegeben, deren Komponenten zufällig aus denselben Komponentensätzen selektiert wurden. Zusätzlich wird eine Kombination aus Punktselektion und lokaler Suche betrachtet, bei der die lokale Suche nach jedem durch die Punktselektion entfernten Versuchspunkt erneut ausgeführt wird. Da jeder entfernte Versuchspunkt andere Versuchspunkte während der Optimierung nicht mehr behindern kann, können auf diese Weise weitere Komponententauschvorgänge durchgeführt und der Versuchsplan so verbessert werden.

In Bild 4.14a ist eine kontinuierliche Verringerung von  $D_{\text{KL}}$  durch die Punktselektion sichtbar, die somit eine gleichförmigere Abdeckung des Eingangsraums beschreibt. Hierbei werden geringfügig bessere Ergebnisse durch die Kombination von Punktselektion und lokaler Suche erzielt. Dieses Verhalten spiegelt sich auch im Vergleich der Modellfehler in Bild 4.14b wider. Während der Fehler des auf dem zufälligen Versuchsplan basierenden Modells kontinuierlich ansteigt, bleibt der Modellfehler bei Verwendung der Punktselektion bis ca.  $N_{\text{Sub}} = 40$  näherungsweise konstant. Auch hier werden durch die Kombination von Punktselektion und lokaler Suche geringfügig bessere Ergebnisse erzielt. Diese Kombination ist daher bei geringen Versuchsumfängen und somit geringer praktischer Bedeutung der Rechenzeit zu bevorzugen.

Die Punktselektion kann auch dazu verwendet werden, um aus einem größeren Satz an Komponenten eines Typs diejenigen Komponenten auszuwählen, die am besten für die Erstellung eines Versuchsplans geeignet sind. Hierbei werden ausschließlich die Merkmale dieses Komponententyps betrachtet, die einen Unterraum des Eingangsraums aufspannen. Wird in diesem Unterraum durch die Punktselektion bereits eine möglichst gleichförmige Datenverteilung erzielt, ermöglicht dies die Erstellung von Gesamtversuchsplänen mit einer hohen Gleichförmigkeit.

## 4.5 Anwendung auf Montageprozesse mit Rotation von Komponenten

Die zuvor beschriebene Methode ist allgemeingültig auf zusammengesetzte Einheiten anwendbar und zielt durch eine systematische Zuordnung verschiedener Einzelkomponenten darauf ab, den Informationsgehalt der experimentell erfassten Daten zu

maximieren. Hierbei liegen jedoch Einschränkungen der realisierbaren Datenpunkte durch die zur Verfügung stehenden Komponenten vor (siehe Bild 4.5b).

Eine Möglichkeit, um diese Einschränkungen zu vermindern und die Vielfalt an möglichen Datenpunkten zu steigern, ist die Einbeziehung anwendungsspezifischer Operationen in den Versuchsplanungsprozess. Diese Operationen verändern die zur Verfügung stehenden Merkmalsausprägungen oder die Randbedingungen durch die gegebenen Komponenten. Damit können weitere Bereiche des Eingangsraums erschlossen werden, was Versuchspläne mit einer höheren Gleichförmigkeit ermöglicht. Beispielsweise können gezielte Manipulationen von Komponenten oder die Montage von Komponenten in unterschiedlichen Positionen zu den anwendungsspezifischen Operationen gezählt werden.

Im Folgenden wird die Rotation von Komponenten bei radialsymmetrischen Baugruppen als eine solche Operation betrachtet. Ein Beispiel für eine radialsymmetrische Baugruppe ist die in Bild 5.11 dargestellte Waschmaschinentrommel. Nachfolgend wird davon ausgegangen, dass unterschiedliche Winkelpositionen vorliegen, an denen Merkmale und Ausgangsgrößen erfasst werden. Bild 4.15 verdeutlicht diese Annahme anhand eines Montageprozesses mit  $I = 2$  Komponenten, die schematisch jeweils als zweidimensionale Projektion in einem eigenen Koordinatensystem dargestellt sind. Bei der Montage werden die Komponenten so aufeinander montiert, dass diese Koordinatensysteme deckungsgleich sind. Jede Komponente wird durch  $m_1 = m_2 = 3$  Merkmale beschrieben, die an  $w = 3$  gleichmäßig verteilten Winkelpositionen entlang des Umfangs der jeweiligen Komponente gemessen werden und bis auf die räumliche Zuordnung dieselbe Messgröße beschreiben. In der montierten Bau-

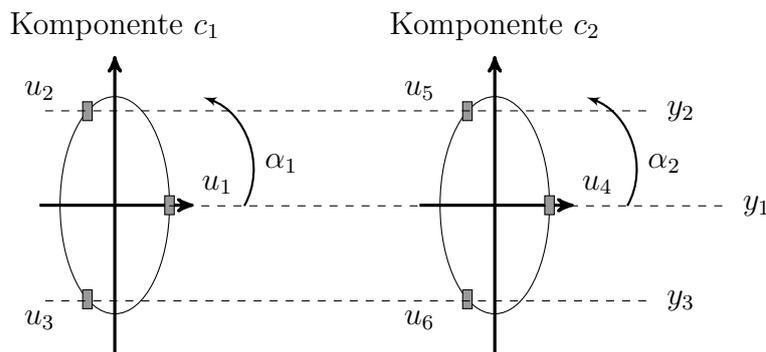


Bild 4.15: Schematische Darstellung einer rotationssymmetrischen Baugruppe aus  $I = 2$  Komponenten, mit den rotatorischen Ausrichtungen  $\alpha_1$  und  $\alpha_2$ . An jeder der  $w = 3$  Messpositionen (gestrichelte Linien) werden Merkmale der Komponenten  $u_1, u_2, \dots, u_6$  und die Ausgangsgrößen  $y_1, y_2, y_3$  gemessen.

gruppe wird an denselben Winkelpositionen die Ausgangsgröße  $y_l$ , mit  $l = 1, 2, \dots, w$  gemessen. Jeder Einzelkomponente wird eine rotatorische Ausrichtung  $\alpha_i(k)$  zugeordnet, die durch den Versuchsplan vorgegeben ist. Nachfolgend werden zwei Ansätze zur Berücksichtigung der Rotation von Komponenten mit unterschiedlichen Auswirkungen auf die Versuchsplanung und das resultierende Prozessmodell vorgestellt.

### 4.5.1 Globaler Versuchsplanungs- und Modellierungsansatz

Für den Montageprozess in Bild 4.15 ist im Allgemeinen die Ausgangsgröße an einer Position der Baugruppe  $\hat{y}_j = f_j(u_1, u_2, u_3, u_4, u_5, u_6)$ , mit  $j = 1, 2, 3$  von allen Merkmalen der Komponenten abhängig und kann jeweils durch ein separates Modell  $f_j(\cdot)$  vorhergesagt werden. Ein Nachteil dieses Ansatzes, bei dem alle gemessenen Merkmale in das Modell einfließen, ist die hohe Dimensionalität des Eingangsraums, die mit einem hohen Bedarf an Datenpunkten einhergeht. Um eine sinnvolle Modellierung mit den geringen Datenmengen zu ermöglichen, die im Rahmen von Versuchen erzielt werden können, ist eine Verdichtung der gemessenen Merkmale zu wenigen aussagekräftigen Merkmalen üblich. In diesem Prozess des *Feature-Engineerings* wird zusätzliches Prozesswissen eingebracht [24]. Nachfolgend werden verdichtete Merkmale der Komponenten betrachtet, denen eine Wirkrichtung zugeordnet werden kann. Durch Rotation einer Komponente kann diese Wirkrichtung und damit der Einfluss der Einzelkomponenten auf die Baugruppe verändert werden. Dieser globale Ansatz wird in Abschnitt 5.3.2 zur Modellierung der Rundlaufabweichungen von Waschmaschinentrommeln verwendet.

#### Einbeziehung des globalen Ansatzes in die Versuchsplanung

Der globale Ansatz basiert darauf, dass jede der  $I$  Komponenten durch  $m_i = 2$  aussagekräftige Merkmale beschrieben wird. Diese Merkmale werden als zweidimensionale Vektoren  $[u_{x,i} \ u_{y,i}]^T$  mit  $i = 1, 2, \dots, I$  dargestellt, um Wirkrichtungen abbilden zu können. Bild 4.16 zeigt ein Beispiel für solche globalen Merkmale, die die radiale Abweichung einer Komponente vom Ursprung eines Referenzkoordinatensystems beschreiben, das beispielsweise in einer Bauteilzeichnung festgelegt wurde. Hierzu wird aus den gemessenen Punkten (Quadrate) ein Ausgleichskreis berechnet. Der Vektor zum Mittelpunkt dieses Kreises dient als globaler Merkmalsvektor.

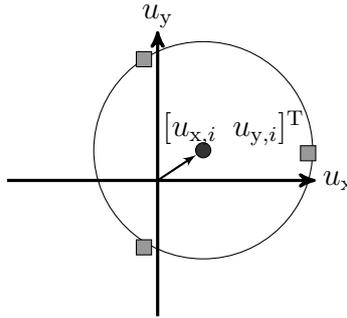


Bild 4.16: Bestimmung der globalen Merkmale einer Einzelkomponente des Typs  $i$ . Der Kreis ergibt sich als Ausgleichskreis aus den Messpunkten (Quadrate). Der Merkmalsvektor charakterisiert die Abweichung des Kreismittelpunkts vom Ursprung.

Durch die Rotation einer Komponente um den Winkel  $\alpha_i(k)$  wird die Richtung des zugehörigen Merkmalsvektors verändert, was mittels der linearen Transformation

$$\begin{bmatrix} \tilde{u}_{x,i}(k) \\ \tilde{u}_{y,i}(k) \end{bmatrix} = \begin{bmatrix} \cos(\alpha_i(k)) & -\sin(\alpha_i(k)) \\ \sin(\alpha_i(k)) & \cos(\alpha_i(k)) \end{bmatrix} \cdot \begin{bmatrix} u_{x,i}(k) \\ u_{y,i}(k) \end{bmatrix} \quad (4.5)$$

beschrieben werden kann. Um optimale Werte für diese Ausrichtungen zu bestimmen, wird der in Abschnitt 4.2.1 beschriebene lokale Suchalgorithmus erweitert. Im Schritt (c) des Algorithmus (siehe Bild A.1) werden nicht nur weitere Versuchspunkte als mögliche Tauschpartner für einen Komponententausch betrachtet, sondern zusätzlich auch alle Punkte, die aus den Rotationen der Komponenten resultieren.

Bild 4.17 stellt exemplarisch den gemessenen Merkmalsvektor  $[-0.2 \ 0.8]^T$  einer Komponente zusammen mit den transformierten Merkmalsvektoren dar, die aus der Rotation der Komponente resultieren. Durch die Betrachtung diskreter Winkelpositionen

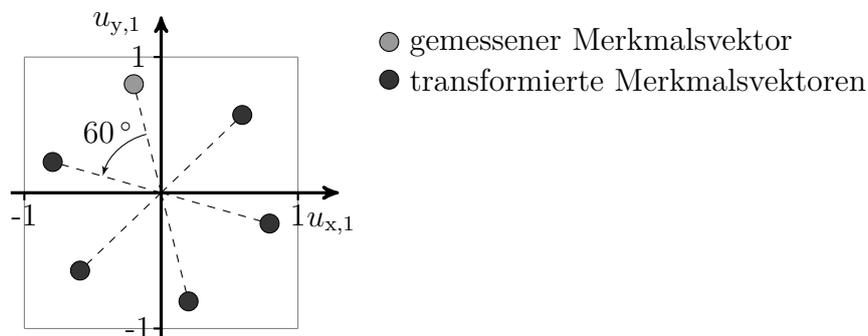


Bild 4.17: Realisierbare Merkmalsvektoren durch Rotation einer Komponente in  $60^\circ$ -Schritten, die als mögliche Tauschpartner bei einem Komponententausch betrachtet werden [115].

wie den dargestellten  $60^\circ$ -Schritten ist hierbei eine endliche Menge an möglichen Tauschpartnern für einen Komponententausch gegeben. Führt ein Tauschvorgang mit einem transformierten Merkmalsvektor zu einer Verbesserung des Optimierungskriteriums, wird neben der Zuordnung der Komponenten zu den Baugruppen auch die rotatorische Ausrichtung der Komponenten im Versuchsplan verändert.

### Demonstration für zwei Komponententypen

Zur Demonstration werden  $N = 20$  Baugruppen mit  $I = 2$  Komponententypen betrachtet, die jeweils durch einen globalen Merkmalsvektor beschrieben werden. Hieraus ergibt sich ein Eingangsraum der Dimensionalität  $p = 4$ , der zur Ermöglichung der Rotation auf  $[-1, 1]^4$  normiert ist. Bild 4.18 stellt die durch beide Komponententypen beeinflussten Unterräume dieses Eingangsraums für zwei Versuchspläne mit und ohne Rotation von Komponenten gegenüber. In (a) ist ein durch Komponententausch optimierter Versuchsplan ohne Rotation von Komponenten dargestellt, sodass  $\alpha_i(k) = 0^\circ \forall i = 1, 2, \dots, I$  und  $k = 1, 2, \dots, N$  gilt. Da die Merkmalsvektoren aller Komponenten eine ähnliche Richtung aufweisen, können die resultierenden Versuchspunkte allein durch Tauschvorgänge von Komponenten nicht weiter über den Eingangsraum verteilt werden. Werden hingegen Rotationen von Komponenten in  $60^\circ$ -Schritten miteinbezogen, ergibt sich die in (b) dargestellte Datenverteilung, die größere Bereiche des Eingangsraums gleichmäßiger abdeckt.

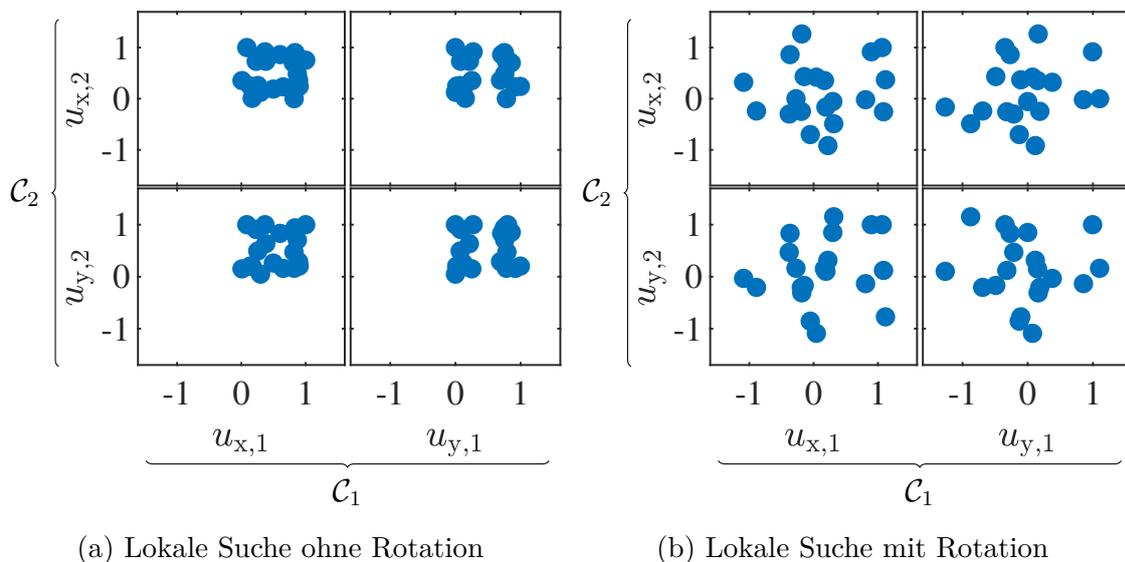


Bild 4.18: Vergleich zweier optimierter Versuchspläne mit und ohne Rotation von Komponenten. Jeder Komponententyp  $\mathcal{C}_1$  und  $\mathcal{C}_2$  wird durch einen globalen Merkmalsvektor beschrieben.



malsausprägungen zu den Positionen aufgehoben und weitere Tauschvorgänge von Komponenten ermöglicht. Weisen Komponenten eine kleine Streuung der Merkmalsausprägungen an einer Position, aber größere Unterschiede zwischen den Positionen auf, kann eine Rotation zu einer deutlichen Verbesserung des Versuchsplans führen. Dies ist beispielsweise häufig bei Komponenten gegeben, die mittels eines Umformwerkzeugs in einer Presse gefertigt werden. Das in Abschnitt 4.2.1 beschriebene Versuchsplanungsverfahren wurde zu diesem Zweck so erweitert, dass neben Tauschvorgängen von Komponenten nacheinander auch alle rotatorischen Ausrichtungen der Einzelkomponenten überprüft werden (vgl. Schritt (e) in Bild A.1). Die Rotation einer Komponente führt zur gleichzeitigen Veränderung von  $w$  lokalen Datenpunkten, die somit alle bei der Auswertung des Optimierungskriteriums berücksichtigt werden müssen. Hierbei verringert die größere Anzahl an gleichzeitig veränderten Punkten die Wahrscheinlichkeit, dass durch eine solche Operation das Optimierungskriterium  $d_{\min}$  verbessert wird. Aus diesem Grund wird die Bedingung für die Durchführung einer Operation aus Gleichung A.3 von „>“ zu „≥“ abgeschwächt, sodass mit

$$\tilde{d}_{\min}^* \geq \tilde{d}_{\min} \quad (4.6)$$

auch Rotationen zugelassen werden, die zu einem gleichbleibenden Wert von  $\tilde{d}_{\min}$  führen. Um trotzdem sicherzustellen, dass ein Tauschvorgang zu einer Verbesserung des Versuchsplans führt, wird mit der zusätzlichen Bedingung

$$\sum_{j=1}^{N \cdot w} \min_{\underline{u}(k), \underline{u}(l) \in \mathcal{D}^*, k \neq l} (d(\underline{u}(k), \underline{u}(l))) > \sum_{j=1}^{N \cdot w} \min_{\underline{u}(k), \underline{u}(l) \in \mathcal{D}(i), k \neq l} (d(\underline{u}(k), \underline{u}(l))) \quad (4.7)$$

die Summe der Nächste-Nachbarn-Distanzen vor und nach einer Operation verglichen. Führt eine Operation zu keiner Verschlechterung von  $\tilde{d}_{\min}$  und steigt die Summe der Nächste-Nachbarn-Distanzen im Versuchsplan an, so wird diese durchgeführt.

### Beispiel zur Optimierung mittels lokalem Ansatz und Rotation

Bild 4.20 stellt beispielhaft zwei Optimierungsschritte für zwei Baugruppen mit  $w = 3$  lokalen Positionen dar. In dem ersten dargestellten Schritt des lokalen Suchalgorithmus wird im Versuchsplan  $\mathcal{D}(1)$  der Punkt  $\underline{u}_2(1)$  (blau) betrachtet, für den der Punkt  $\underline{u}_1(1)$  (orange) als möglicher Tauschpartner für das Merkmal  $\tilde{u}_1$  ausgewählt wurde. Beide Punkte gehören zur selben Baugruppe  $\mathcal{P}(1)$ , sodass ausschließlich eine Rotation der Komponente erfolgt, die die  $\tilde{u}_1$  Werte dieser Punkte vorgibt. Dem betrachteten Punkt  $\underline{u}_2(1)$  wird hierbei die Merkmalsausprägung des Tauschpartners

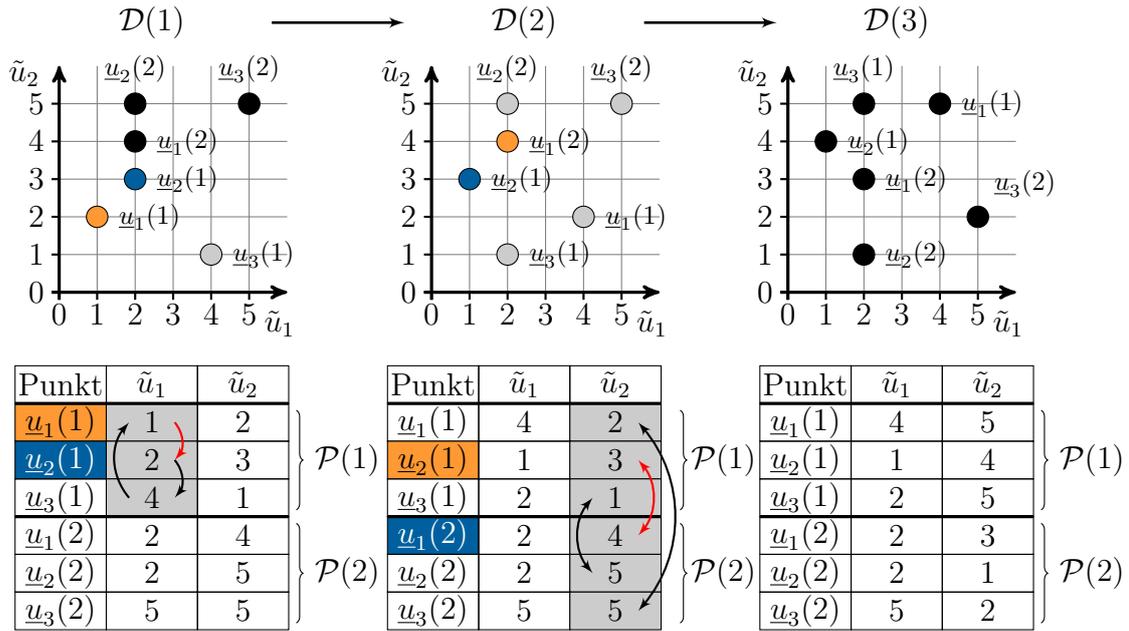


Bild 4.20: Beispiel für zwei Optimierungsschritte unter Verwendung des lokalen Ansatzes bei der Rotation von Komponenten [116]. Zwei Baugruppen  $\mathcal{P}(1)$  und  $\mathcal{P}(2)$  mit  $w = 3$  Positionen sind gegeben, sodass jede Baugruppe drei Datenpunkte definiert. Im Vergleich zum Versuchsplan  $\mathcal{D}(1)$  ist in  $\mathcal{D}(2)$  eine Komponente rotiert. In  $\mathcal{D}(3)$  wurde ein Tauschvorgang mit gleichzeitiger Rotation durchgeführt.

$\underline{u}_1(1)$  zugeordnet (roter Pfeil). Alle weiteren Tauschoperationen der beteiligten Punkte werden entsprechend der Rotation ausgeführt.

Da für die dargestellte Operation beide Bedingungen gemäß Gleichung (4.6) und Gleichung (4.7) erfüllt sind, wird die nächste Iteration mit dem Versuchsplan  $\mathcal{D}(2)$  fortgesetzt. In  $\mathcal{D}(2)$  wird  $\underline{u}_1(2)$  (blau) betrachtet und  $\underline{u}_2(1)$  (orange) als ein geeigneter Tauschpartner für das Merkmal  $\tilde{u}_2$  identifiziert. Da diese Punkte zu unterschiedlichen Baugruppen gehören, wird ein Komponententausch mit gleichzeitiger Rotation durchgeführt. Die Rotation erfolgt so, dass ein direkter Tausch der Merkmalsausprägung von  $\tilde{u}_2$  zwischen den ausgewählten Punkten  $\underline{u}_2(1)$  und  $\underline{u}_1(2)$  stattfindet (roter Pfeil). Alle weiteren lokalen Datenpunkte der betroffenen Baugruppen werden zugleich durch die Rotation verändert. Auch für diese Operation sind die Bedingungen gemäß Gleichung (4.6) und Gleichung (4.7) erfüllt. Im Vergleich zwischen  $\mathcal{D}(1)$  und  $\mathcal{D}(3)$  zeigt sich, dass durch die Rotation von Komponenten in Verbindung mit dem lokalen Ansatz die minimale Punktdistanz  $d_{\min}$  gesteigert und so eine höhere Gleichförmigkeit des Versuchsplans erzielt werden kann.

Sowohl die Steigerung der Datenmenge als auch die Reduzierung der Dimensionalität durch den lokalen Ansatz sind vorteilhaft für die Modellbildung. Somit werden die typischen Probleme in der Versuchsplanung für zusammengesetzte Einheiten von kleinen Stichprobenumfängen und einer großen Anzahl an Merkmalen pro Komponente vermindert. Allerdings ist es hierdurch nicht möglich, Interaktionen zwischen den verschiedenen Positionen abzubilden, was die Anwendbarkeit dieses Ansatzes einschränkt.

## 4.6 Zusammenfassung

Um Versuche mit zusammengesetzten Einheiten mit überschaubarem Aufwand und Kosten praktisch umsetzen zu können, wurde in diesem Kapitel eine Versuchsplanungsmethode vorgestellt, die auf gegebene Sätze an Komponenten zurückgreift. Ein resultierender Versuchsplan beschreibt die Zusammenstellung der Komponenten zu Einheiten und zielt darauf ab, raumfüllende Datenverteilungen zu erzeugen. Hierbei müssen im Gegensatz zu klassischen Versuchsplanungsmethoden zusätzliche Randbedingungen hinsichtlich der realisierbaren Merkmalsausprägungen, dem Erhalt fester Kombinationen von Merkmalsausprägungen und der einmaligen Verwendung der Komponenten berücksichtigt werden. Zur Erstellung solcher Versuchspläne unter Berücksichtigung dieser Randbedingungen wurden mit einem lokalen Suchalgorithmus und einem genetischen Algorithmus zwei verschiedene Optimierungsverfahren betrachtet. Im Vergleich erzielte der lokale Suchalgorithmus bessere Ergebnisse und benötigte zugleich keine Anpassung von Hyperparametern.

Durch ein geändertes Optimierungskriterium wurden mit demselben lokalen Suchalgorithmus möglichst ähnliche Versuchspunkte erzeugt. Es wurde gezeigt, dass eine geringe Anzahl solcher ähnlichen Versuchspunkte in Verbindung mit einem Gaußprozessmodell zu einer verbesserten Schätzung von Rauscheinflüssen beitragen kann.

Als Möglichkeit zur Steigerung der Gleichförmigkeit von Versuchsplänen und zur Reduzierung des Versuchsaufwands wurde ein Verfahren zur Punktselektion beschrieben. Hiermit ist es möglich, sequentiell Versuchspunkte zu entfernen, die hauptsächlich redundante Informationen tragen und nicht weiter optimiert werden können. An einem Beispiel wurde gezeigt, dass bei gegebenen Datenverteilungen unter Umständen viele Versuchspunkte entfernt werden können und somit nicht vermessen werden müssen, ohne zu einer signifikanten Verschlechterung der Modellgüte zu führen.

---

Neben der gezielten Zusammenstellung von Komponenten können anwendungsspezifische Operationen zu einer Verbesserung eines Versuchsplans genutzt werden. Die Rotation von Komponenten in einer radialsymmetrischen Baugruppe wurde als eine solche Operation betrachtet. Hierfür wurden zwei verschiedene Ansätze vorgestellt, die unterschiedliche Annahmen an den zu modellierenden Prozess stellen. Für beide Ansätze wurde gezeigt, wie durch die Rotation von Komponenten eine gleichförmigere Datenverteilung erzielt werden kann.



## 5 Anwendungen

In diesem Kapitel werden verschiedene praktische Anwendungen der in dieser Arbeit entwickelten Versuchsplanungsverfahren beschrieben. In Abschnitt 5.1 wird die inkrementelle Modellbildung an einem Labormodell einer Heizstrecke demonstriert, wobei verschiedene inkrementelle Versuchspläne anhand dieses realitätsnahen Prozesses verglichen werden. Abschnitt 5.2 präsentiert die systematische Vermessung und Modellierung eines Bitumenofens unter Verwendung der inkrementellen Modellbildung. Auf Basis des erstellten Prozessmodells wird eine optimale Steuerung realisiert und im regulären Betrieb validiert. In Abschnitt 5.3 werden Versuche für Montageprozesse von Waschmaschinentrommeln beschrieben. Hierbei werden mit einem lokalen und einem globalen Modellansatz zwei verschiedene Strategien mit unterschiedlichen Auswirkungen auf die Versuchsplanung und die Modellbildung angewandt. Die entwickelten Modelle können direkt dazu eingesetzt werden, die im Montageprozess erzielte Qualität zu steigern.

### 5.1 Inkrementelle Modellbildung einer Heizstrecke

Die in diesem Abschnitt betrachtete Heizstrecke ist ein Labormodell der Firma ELWE Technik, das viele Eigenschaften realer industrieller Prozesse aufweist und mittels der in Abschnitt 3 beschriebenen inkrementellen Modellbildung modelliert wird. Als Labormodell ermöglicht die Heizstrecke im Vergleich zu produktiv genutzten Prozessen eine sichere Durchführung und einen größeren Umfang von Versuchen. Somit ist eine Vermessung mehrerer Versuchspläne praktikabel, auf deren Basis verschiedene Modelle erstellt und verglichen werden können.

Die Heizstrecke mit deren wichtigsten Komponenten ist in Bild 5.1 dargestellt. Sie besteht aus einem Lüfter, der mit variabler Leistung  $P_L$  betrieben werden kann. Die geförderte Luft wird durch eine Heizung mit der Heizleistung  $P_H$  erwärmt und gelangt in ein Rohr, in dem die Lufttemperatur  $\vartheta_L$  und der Luftmassenstrom  $\dot{m}_L$  gemessen

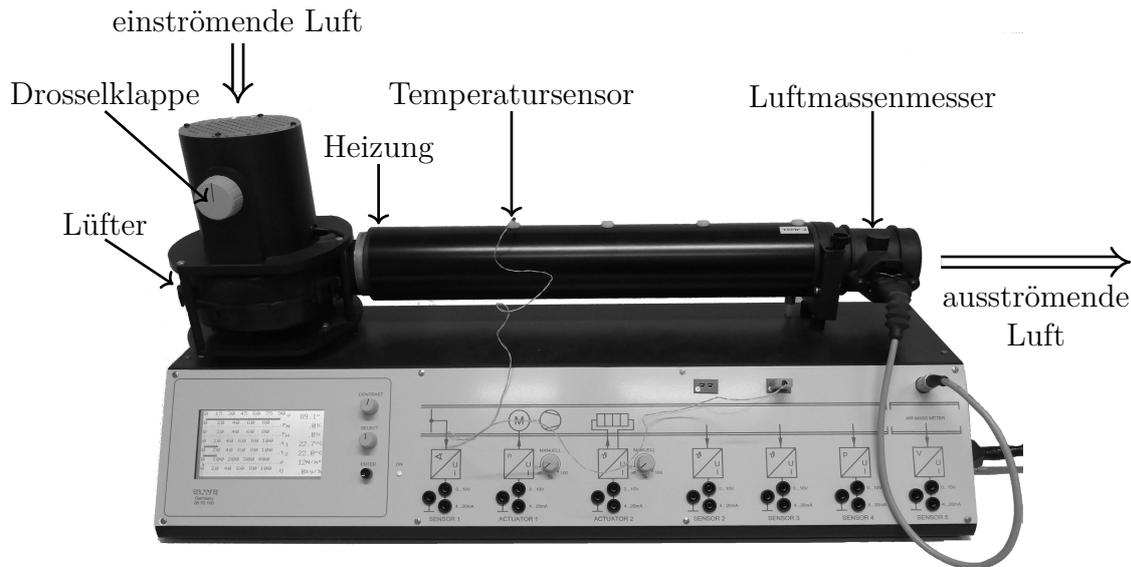


Bild 5.1: Betrachtetes Labormodell einer Heizstrecke [113].

werden. Zusätzlich ist eine manuell zu bedienende Drosselklappe am Lufteinlass verbaut, deren Winkelposition  $\beta$  erfasst wird. Durch eine Kopplung mit *Matlab-Simulink* ist eine automatische Erfassung aller Messdaten sowie eine Vorgabe der Eingangsgrößen  $P_L$  und  $P_H$  möglich. Abgesehen von der Variation der Drosselklappenposition können somit automatische Versuche durchgeführt werden. Um eine Überhitzung der Heizstrecke zu vermeiden, ist eine maximale Lufttemperatur vorgegeben. Liegt die Lufttemperatur mit  $\vartheta_L > 100\text{ }^\circ\text{C}$  außerhalb des zulässigen Bereichs, erfolgt eine Notabschaltung der Heizstrecke. Diese Begrenzung muss bei der Versuchsplanung berücksichtigt werden und führt dazu, dass nicht alle Kombinationen der Eingangsgrößen zulässig sind. So kann beispielsweise eine hohe Heizleistung bei niedriger Lüfterleistung zu einer unzulässig hohen Lufttemperatur führen.

Zusammenfassend besitzt die Heizstrecke folgende Eigenschaften, die in der nachfolgenden inkrementellen Modellbildung berücksichtigt werden:

- Es sind drei Eingangsgrößen  $P_L$ ,  $P_H$  und  $\beta$  gegeben, die das Prozessverhalten unterschiedlich stark beeinflussen und miteinander interagieren.
- Es sind zwei Ausgangsgrößen  $\vartheta_L$  und  $\dot{m}_L$  gegeben, von denen eine einen vorgegebenen Grenzwert nicht überschreiten darf.
- Der Prozess weist dynamisches Verhalten auf, soll aber statisch modelliert werden. Daher erfolgt die Messung der Ausgangsgrößen erst, wenn ein stationärer Zustand erreicht ist.

### 5.1.1 Anwendung der inkrementellen Modellbildung

Die inkrementelle Modellbildung der Heizstrecke zielt auf die Lufttemperatur  $\vartheta_L$  als zu modellierende Ausgangsgröße ab, da diese von allen drei Eingangsgrößen signifikant beeinflusst wird. Der Luftmassenstrom  $\dot{m}_L$  als zweite mögliche Ausgangsgröße unterliegt hingegen vernachlässigbaren Einflüssen durch die Heizleistung.

#### Rangfolge und Wertebereiche der Eingangsgrößen

Auf Grundlage physikalischer Überlegungen erfolgt die Festlegung einer Rangfolge der Eingangsgrößen, woraus sich der Eingangsgrößenvektor  $\underline{u} = [P_L \ P_H \ \beta]^T$  ergibt. In der gewählten Rangfolge steht die Lüfterleistung an erster Stelle, da diese gegenüber der Heizleistung mit einer kleineren Zeitkonstante auf die Ausgangsgröße wirkt. Bei ausschließlicher Variation dieser Größe wird ein stationärer Zustand schneller erreicht, was die Versuchsdauer im ersten Schritt der inkrementellen Modellbildung verringert. Die inkrementelle Modellbildung erfasst das Prozessverhalten ausgehend von einem Arbeitspunkt, der an der Heizstrecke zu  $\underline{\bar{u}} = [43\% \ 57\% \ 44^\circ]^T$  gewählt wird. Um ein Überschreiten der zulässigen Lufttemperatur während der Versuche zu vermeiden, erfolgt eine Beschränkung der Wertebereiche der Eingangsgrößen auf  $P_L \in [15\%, 85\%]$ ,  $P_H \in [15\%, 85\%]$  und  $\beta \in [20^\circ, 80^\circ]$ . Hierdurch ist wie in Kapitel 3, ein quaderförmiger Eingangsraum gegeben. Eine praktische Anwendung der inkrementellen Modellbildung in einem aufgrund von Randbedingungen nicht quaderförmigen Eingangsraum ist in Abschnitt 5.2 beschrieben.

#### Versuchspläne und Modelle

Die Heizstrecke wird mit verschiedenen inkrementellen Modellen modelliert, wobei Polynommodelle, lokale Modellnetze (LMN) und Gaußprozessmodelle (GPM) zum Einsatz kommen. Um den Nutzen der inkrementellen Modellbildung aufzuzeigen und mit einem klassischen Vorgehen zur experimentellen Modellbildung zu vergleichen, werden folgende Messungen an der Heizstrecke durchgeführt:

1. Ein inkrementeller Versuchsplan für Polynommodelle, LMN und GPM,
2. Zwei inkrementelle Versuchspläne mit Transformation für LMN,
3. Ein optimierter Latin Hypercube (LHC) für Polynommodelle, LMN und GPM,
4. Zwei Testdatensätze mit unterschiedlicher Streuung um den Arbeitspunkt,
5. Ein raumfüllender Testdatensatz auf Basis einer Sobol-Sequenz.

Alle inkrementellen Modelle werden jeweils in  $p = 3$  Schritten erstellt, wobei  $\underline{N} = [21 \ 20 \ 20]^T$  Versuchspunkte vermessen werden. Auf Basis des Versuchsplans in 1. werden mit einem Polynommodell, einem LMN und zwei GPM (GPM-DP und GPM-NS) vier verschiedene inkrementelle Modelle erstellt. Die Versuchspläne in 2. dienen zur Erstellung zweier weiterer inkrementeller Modelle mit LMN. Hierbei werden die LHC-Stufen mittels einer Transformation angepasst, um die Dichte an Versuchspunkten in nichtlinearen Bereichen des Eingangsraums zu steigern. Die Transformation erfolgt auf Grundlage der Partitionierung der LMN in den vorherigen Schritten (siehe Abschnitt 3.3.2), wobei mit  $r = 0$  und  $r = -0.5$  zwei unterschiedliche Stärken dieser Transformation betrachtet werden. Als ein klassisches Vorgehen zur experimentellen Modellbildung wird mit dem Versuchsplan in 3. ein optimierter LHC für alle  $p = 3$  Eingangsgrößen vermessen. Dieser LHC besitzt dieselbe Anzahl an Punkten  $\|\underline{N}\|_1 = N_1 + N_2 + N_3 = 61$  wie ein inkrementeller Versuchsplan nach  $p = 3$  Schritten und wird zur Erstellung von drei Vergleichsmodellen aus unterschiedlichen Modellklassen (Polynom, LMN und GPM) verwendet. Für alle Modelle werden die in Kapitel 3 beschriebenen Standardparameter verwendet. Hierzu zählt ein Grad von  $m = 2$  für alle Polynommodelle und eine Fehlerschwelle von  $\epsilon = 0.5^\circ\text{C}$  als Abbruchkriterium für den LOLIMOT-Algorithmus.

Die Bewertung der Modelle erfolgt anhand von Testdatensätzen, bei denen der vorgegebene Arbeitspunkt im Fokus steht. Diese Testdatensätze werden in 4. gemäß Gleichung 3.2 aus um den Arbeitspunkt  $\bar{u}$  zentrierten Normalverteilungen erstellt. Mit  $\sigma_{\text{test}} = 0.05$  und  $\sigma_{\text{test}} = 0.2$  werden unterschiedliche Standardabweichungen für jeden Testdatensatz verwendet, wobei jeweils  $N_{\text{test}} = 20$  Datenpunkte vermessen werden. Zusätzlich wird mit dem Versuchsplan in 5. ein Testdatensatz mit  $N_{\text{test}} = 20$  Datenpunkten auf Basis einer Sobol-Sequenz erstellt. Dieser raumfüllende Versuchsplan dient zur Beurteilung der Modellgüte im gesamten Eingangsraum.

### Versuchsdurchführung

Für jeden Versuchsplan bzw. Schritt in der inkrementellen Modellbildung werden nacheinander die  $N$  Versuchspunkte  $\underline{u}(k)$  mit  $k = 1, 2, \dots, N$  vermessen. Hierzu werden die durch den jeweiligen Punkt vorgegebenen Kombinationen an Eingangsgrößen an der Heizstrecke eingestellt. Da die inkrementelle Modellbildung darauf abzielt, das statische Prozessverhalten abzubilden, werden diese Eingangsgrößen für eine Dauer von 25 s konstant gehalten, sodass ein stationärer Zustand der Lufttemperatur erreicht wird. Der statische Prozessausgang  $\vartheta_L(k)$  ergibt sich zur Vermin-

derung von Messrauschen als arithmetischer Mittelwert über die letzten 4 s dieser Zeitspanne, wobei die zugrunde liegenden Messwerte mit einer Frequenz von 10 Hz abgetastet werden. In der Versuchsdurchführung an der Heizstrecke wird ein Vorteil der inkrementellen Modellbildung gegenüber klassischen Versuchsplanungsverfahren deutlich. Da die Drosselklappenposition, die mit einem hohen Stellaufwand manuell eingestellt werden muss, erst im Schritt  $i = 3$  einbezogen wird, sind vollautomatische Versuche in den Schritten  $i = 1$  und  $i = 2$  möglich. Insgesamt sind somit 21 manuelle Einstellvorgänge der Drosselklappe nötig, während die Position bei dem optimierten LHC mit derselben Anzahl an Versuchspunkten 61 Mal variiert werden muss.

### 5.1.2 Ergebnisse

Bild 5.2 zeigt exemplarisch die gemessene Lufttemperatur  $\vartheta_L$  für die Versuchspunkte in den Schritten  $i = 1$  und  $i = 2$  sowie die Modellausgänge für zwei verschiedene inkrementelle Modelle mit der Drosselklappenposition  $\beta = 44^\circ$ . Beide Modelle sind in der Lage, das Prozessverhalten abzubilden, wobei das GPM in (b) insgesamt einen glatteren Verlauf der Ausgangsgröße erzielt. Beim LMN in (a) wurde mit  $r = 0$  eine Transformation der LHC-Stufen durchgeführt. Die durch diese Transformation an die Nichtlinearität angepasste Datenverteilung ist im Bild sichtbar und führt im Schritt  $i = 2$  zu einer höheren Dichte an Versuchspunkten im Bereich  $P_L < 50\%$ .

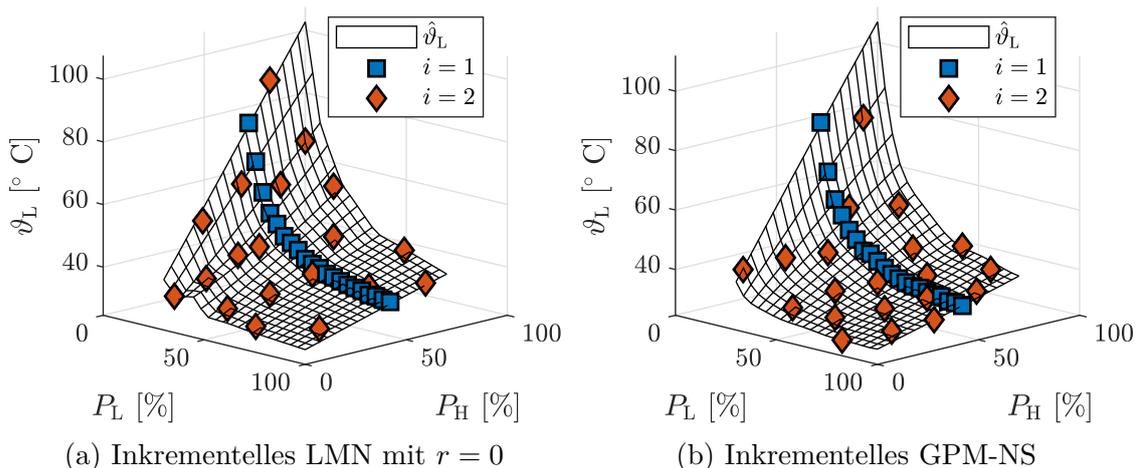
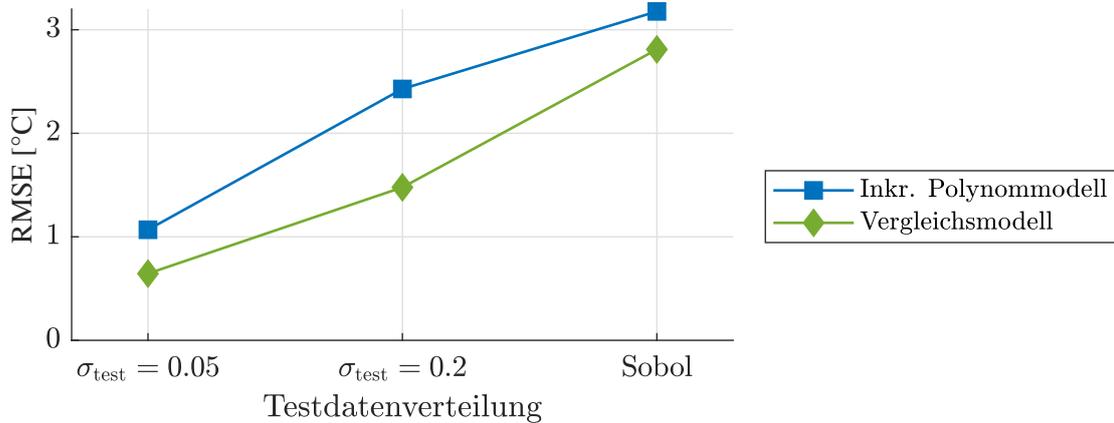


Bild 5.2: Messwerte der Lufttemperatur  $\vartheta_L$  für zwei inkrementelle Versuchspläne in den Schritten  $i = 1$  und  $i = 2$  mit  $\beta = 44^\circ$ . Zusätzlich sind die Modellausgänge  $\hat{\vartheta}_L$  von zwei auf Basis der jeweiligen Versuchsdaten erstellten inkrementellen Modelle dargestellt.

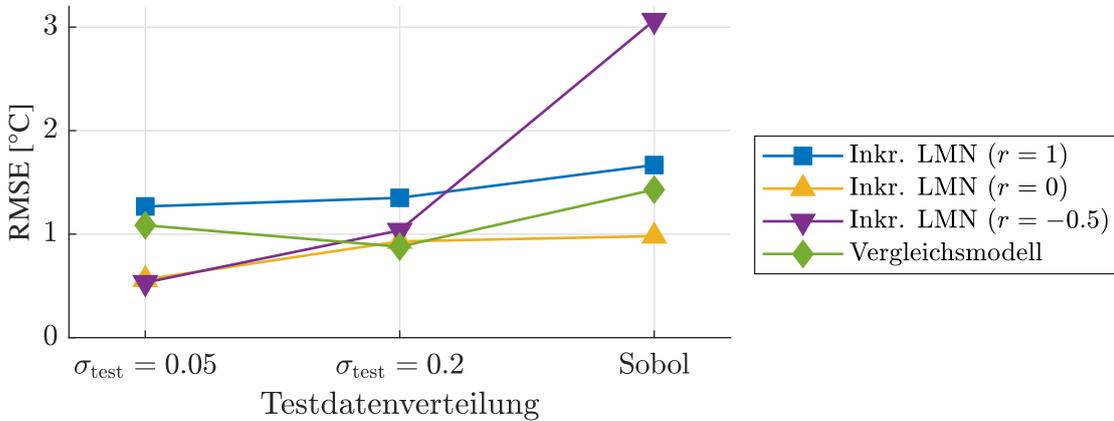
Die verschiedenen inkrementellen Modelle und die zugehörigen Vergleichsmodelle werden auf Basis der Testdatensätze miteinander verglichen. Die erzielten Ergebnisse sind in Bild 5.3 nach den einzelnen Modellklassen geordnet dargestellt. Bei allen inkrementellen Modellen aller Modellklassen ist eine Zunahme des Fehlers mit steigender Raumabdeckung der Testdaten gegeben. So erreichen die inkrementellen Modelle ihre höchste Güte in der Nähe des Arbeitspunkts mit  $\sigma_{\text{test}} = 0.05$ , während jeweils ein größerer Fehler bei Verwendung der Sobol-Sequenz als Testdatensatz gegeben ist. Die Vergleichsmodelle sind durch die größere Anzahl an gleichmäßiger verteilten Datenpunkten in der Lage, das Prozessverhalten für den gesamten Eingangsraum besser abzubilden. Hierdurch weisen die Vergleichsmodelle abgesehen vom inkrementellen LMN mit  $r = 0$  jeweils den geringsten Fehler bei der Sobol-Sequenz als Testdatensatz auf.

Das polynomiale Vergleichsmodell in (a) erzielt bei allen Testdatensätzen einen geringeren Fehler als das inkrementelle Polynommodell. Dies kann durch die unzureichende Flexibilität der Polynommodelle erklärt werden, die bei dem inkrementellen Modell bereits in den ersten Schritten zu großen Modellfehlern führt. Diese Fehler können in weiteren Schritten nicht mehr korrigiert werden. Bei den LMN in (b) ist mit zunehmender Transformationsstärke von  $r = 1$  bis  $r = -0.5$  eine Steigerung der Modellgüte in der Nähe des Arbeitspunkts sichtbar. Durch die Transformation können mehr lokale Modelle in nichtlinearen Bereichen des Eingangsraums erstellt werden, was bei dem LMN mit  $r = 0$  zu einer insgesamt gesteigerten Modellgüte führt, die bei der Sobol-Sequenz als Testdatensatz sogar das Vergleichsmodell übertrifft. Wird die Transformation mit  $r = -0.5$  zu stark gewählt, steigt der Fehler im gesamten Eingangsraum durch die ungleichmäßige Verteilung der Trainingsdaten an. Für die inkrementellen GPM in (c) sind für alle Testdatensätze bessere Ergebnisse von GPM-NS gegenüber GPM-DP zu erkennen. Beide inkrementelle Modelle bieten für  $\sigma_{\text{test}} = 0.05$  eine höhere Modellgüte als das Vergleichsmodell, während für  $\sigma_{\text{test}} = 0.2$  und die Sobol-Sequenz das Vergleichsmodell die besten Ergebnisse erzielt.

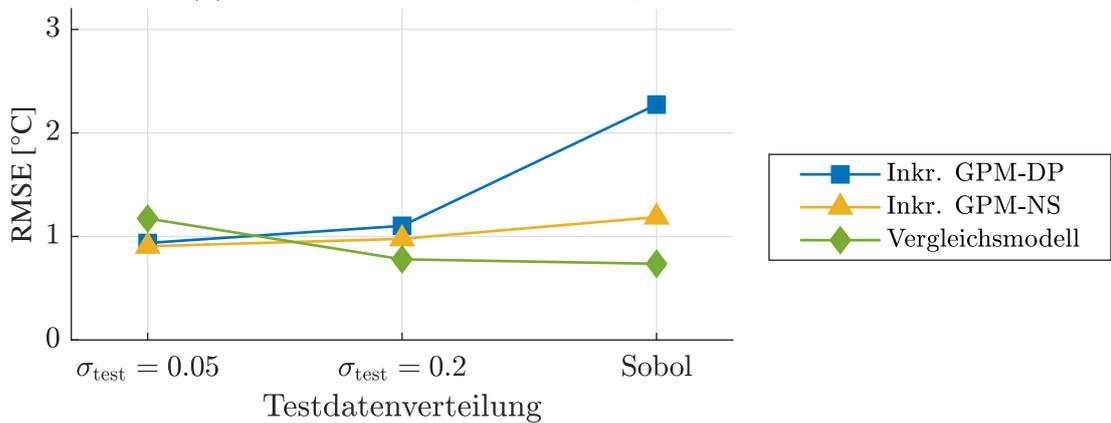
Im Vergleich der verschiedenen Modellklassen sind ähnliche Fehlerwerte für  $\sigma_{\text{test}} = 0.05$  zu erkennen, was verdeutlicht, dass alle Modelle in der Lage sind, die funktionalen Zusammenhänge in einem Bereich um den Arbeitspunkt zu approximieren. Den insgesamt geringsten Fehler erzielt hier das LMN mit der stärksten Transformation  $r = -0.5$ . Für  $\sigma_{\text{test}} = 0.2$  und die Sobol-Sequenz können geringere Fehler mit den flexibleren Modellklassen LMN und GPM erzielt werden als mit den Polynommodellen. Bei Betrachtung der Sobol-Sequenz als Testdatensatz ist das LMN mit



(a) Inkrementelles Modell und Vergleichsmodell mit polynomialer Struktur



(b) Inkrementelles Modell und Vergleichsmodell mit LMN



(c) Inkrementelles Modell und Vergleichsmodell mit GPM

Bild 5.3: Vergleich verschiedener Modelle der Heizstrecke anhand des RMSE unter Verwendung verschiedener Testdatensätze. Diese Testdatensätze decken von links nach rechts einen zunehmenden Anteil des Eingangsraums mit Datenpunkten ab.

$r = 0$  das beste inkrementelle Modell mit einem RMSE von  $0.98^\circ\text{C}$  und das GPM als Vergleichsmodell insgesamt das beste Modell mit einem RMSE von  $0.74^\circ\text{C}$ .

Zusammenfassend zeigt sich anhand dieser Anwendung, dass die inkrementelle Modellbildung in der Lage ist, eine vergleichbare bzw. in der Nähe des Arbeitspunkts sogar höhere Modellgüte zu erzielen als ein klassisches Vorgehen zur experimentellen Modellbildung. Zugleich werden Vorteile bei der Versuchsdurchführung durch das schrittweise Vorgehen deutlich, die bei dieser Anwendung eine Verringerung des Stellaufwands bewirken. Außerdem wurde der praktische Nutzen der Transformation von Versuchsplänen auf Basis von Informationen aus den LMN der vorherigen Schritte anhand dieser Anwendung demonstriert.

## 5.2 Modellbasierte Optimierung eines Bitumenofens

In diesem Abschnitt wird ein industrieller Ofen betrachtet, der bei der Firma Miele & Cie. KG zum Aufschmelzen von Bitumenmatten auf Laugenbehälter dient. In einer Waschmaschine rotiert die Trommel mit der zu waschenden Wäsche innerhalb eines solchen Laugenbehälters, wobei die Bitumenmatte zu einer Geräuschreduktion während des Waschvorgangs führt. Mehrere Laugenbehälter mit aufgelegten Bitumenmatten sind in Bild 5.4 am Eingang des Ofens zu sehen. Im Ofen werden die Laugenbehälter und Bitumenmatten mittels Infrarotstrahlung erwärmt, wobei die

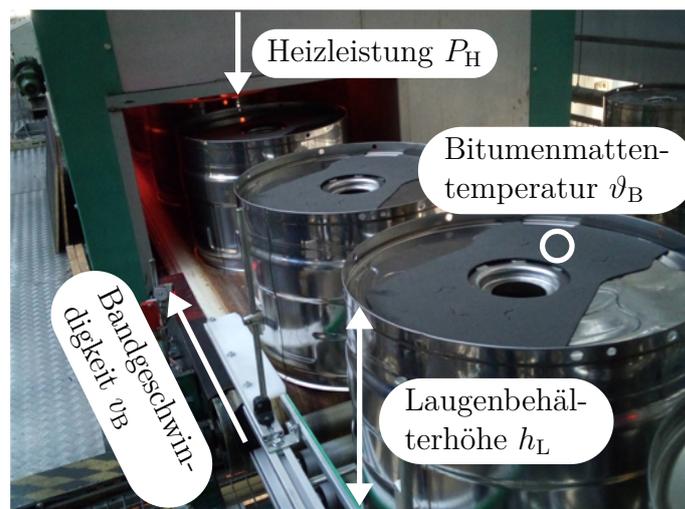


Bild 5.4: Laugenbehälter mit aufgelegten Bitumenmatten am Eingang des Bitumenofens [114]. Die betrachteten Eingangsgrößen und die Bitumenmattentemperatur  $v_B$  als Qualitätsgröße sind kenntlich gemacht.

Heizleistung  $P_H$  in diskreten Stufen variiert werden kann. Die Bandgeschwindigkeit  $v_B$ , mit der sich die Laugenbehälter durch den Ofen bewegen, bestimmt die Verweilzeit unter den Infrarotstrahlern. Als weitere Eingangsgröße wird die Laugenbehälterhöhe  $h_L$  miteinbezogen, die aufgrund der Abstrahlcharakteristik der Infrarotstrahler die Erwärmung der Bauteile beeinflusst. Als Qualitätsgröße für den Prozess dient die Bitumenmattentemperatur  $\vartheta_B$ , die nach dem Durchlaufen des Ofens kontaktlos auf jeder Bitumenmatte gemessen wird. Für diese Temperatur sind Grenzwerte einzuhalten, um ein Ablösen oder Überhitzen der Bitumenmatte zu vermeiden.

Ziel ist es, eine vorgegebene Bitumenmattentemperatur im regulären Betrieb des Ofens möglichst exakt einzuhalten. Hierzu wird eine Optimierungsstrategie entwickelt, die unter Berücksichtigung von zusätzlichen Produktionsvorgaben optimale Werte für die Eingangsgrößen des Bitumenofens vorgibt. In einem ersten Schritt wird der Bitumenofen mittels der ILHAD-Methodik (siehe Kapitel 3) vermessen und datenbasiert modelliert, um ein Prozessmodell zur Vorhersage der Bitumenmattentemperatur zu erhalten. Am Modell werden anschließend mittels eines Optimierungsverfahrens Betriebsparameter ermittelt, wobei Produktionsvorgaben berücksichtigt werden. Eine solche Produktionsvorgabe ist der Typ des verarbeiteten Laugenbehälters. Für jeden Laugenbehältertyp mit spezifischer Höhe werden somit optimale Werte der Bandgeschwindigkeit und Heizleistung bestimmt. Neben der Sicherstellung einer gleichbleibenden Produktqualität wird ein minimaler elektrischer Energiebedarf des Ofens als weiteres Optimierungsziel berücksichtigt. Die Implementierung der Optimierungsstrategie in der Anlagensteuerung ermöglicht eine fortlaufende Anpassung der Betriebsparameter an variierende Betriebsbedingungen. Erste Ergebnisse der Modellbildung des Bitumenofens wurden in [114] veröffentlicht. Weitere Details zu dieser Anwendung sind in [82] zu finden.

### 5.2.1 Anwendung der inkrementellen Modellbildung

Eine Fertigungsanlage wie der betrachtete Bitumenofen verfügt über zahlreiche Parameter, die das Prozessverhalten beeinflussen. Viele dieser Parameter wie die Geometrie des Ofenraums wurden bei der Konstruktion festgelegt und können nur mit großem Aufwand verändert werden. Weitere Parameter wie die Dicke der Bitumenmatten unterliegen genauen Vorgaben und können somit als konstant angenommen werden. Für die Modellierung des Bitumenofens wurden diejenigen Prozessparameter als Eingangsgrößen ausgewählt, die gemäß dem vorhandenen Prozesswissen den

größten Einfluss auf den Prozess haben und einfach auf einen durch den Versuchsplan vorgegebenen Wert eingestellt werden können. Diese Größen sind in Tabelle 5.1 aufgelistet und mit vorhandenem Prozesswissen bewertet worden. Die Bandgeschwindigkeit und die Heizleistung wurden als die bedeutendsten Eingangsgrößen identifiziert, wobei die Bandgeschwindigkeit in der Rangfolge aufgrund der besseren Einstellbarkeit priorisiert wird. Die Heizleistung weist für die Versuchsdurchführung eine schlechtere Einstellbarkeit auf, da sich diese aufgrund von Aufheiz- bzw. Abkühlvorgängen der Heizstrahler verzögert auf den Prozess auswirkt. Die Laugenbehälterhöhe kann durch manuelles Zuführen verschiedener Bauteiltypen während der Versuche eingestellt werden, besitzt aber eine geringere Bedeutung, sodass diese an dritter Stelle in der Rangfolge steht. Somit ergibt sich für die inkrementelle Modellbildung der Eingangsgrößenvektor  $\underline{u} = [v_B \ P_H \ h_L]^T$ , mit dem die Bitumenmattentemperatur  $\hat{\vartheta}_{B,\text{inkr}} = f(\underline{u})$  vorhergesagt werden soll. Als Arbeitspunkt wird der in der Vergangenheit häufigste Betriebspunkt  $\underline{\bar{u}} = [0.058 \text{ m/s} \ 56 \text{ kW} \ 0.412 \text{ m}]^T$  gewählt.

Die Ofentemperatur  $\vartheta_O$  wird nicht in das inkrementelle Modell einbezogen. Sie weist ein dynamisches Verhalten auf und wird durch weitere Eingangsgrößen beeinflusst, wobei der größte Einfluss von der Heizleistung ausgeht. Da die Versuche mit variabler Heizleistung stattfinden, kann die Ofentemperatur folglich nicht konstant gehalten werden. Stattdessen wird ein polynomiales Korrekturmodell mit Wechselwirkung

$$\Delta\hat{\vartheta}_B = g(v_B, \vartheta_O) = b_0 + b_1v_B + b_2\vartheta_O + b_3v_B\vartheta_O \quad (5.1)$$

verwendet, das den Einfluss der Ofentemperatur im für den Betrieb relevanten Temperaturbereich näherungsweise kompensiert. Da der Einfluss der Ofentemperatur von der Verweilzeit der Laugenbehälter im Ofen abhängt, wird in diesem Modell auch die Bandgeschwindigkeit berücksichtigt. Die Parameter  $b_0, b_1, b_2$  und  $b_3$  wurden mit der Methode der kleinsten Fehlerquadrate (siehe Abschnitt 2.2.1) auf Grundlage eines gesonderten vollfaktoriellen Versuchsplans (siehe Abschnitt 2.4.2) mit vier Versuchs-

Tabelle 5.1: Betrachtete Eingangsgrößen am Bitumenofen und Kriterien zur Festlegung einer Rangfolge.

Bezeichnung	Stufen	Bedeutung	Einstellbarkeit	Rangfolge
Bandgeschwindigkeit $v_B$	keine	+	+	1
Heizleistung $P_H$	5	+	0	2
Laugenbehälterhöhe $h_L$	5	0	+	3
Ofentemperatur $\vartheta_O$	keine	0	-	/

punkten geschätzt. Aus dem inkrementellen Modell und dem Korrekturmodell ergibt sich die vorhergesagte Bitumenmattentemperatur  $\hat{v}_B = \hat{v}_{B,\text{inkr}} + \Delta\hat{v}_B$ .

### Inkrementelle Versuchsplanung unter Berücksichtigung von Randbedingungen

Am Bitumenofen existieren Kombinationen von Eingangsgrößen, die aufgrund physikalischer Überlegungen ausgeschlossen werden können. So ist bei einer hohen Bandgeschwindigkeit und geringen Heizleistung mit einer zu niedrigen Bitumenmattentemperatur zu rechnen. Eine Vermessung dieses Bereichs hätte kaum praktischen Nutzen, würde aber den Versuchsaufwand erhöhen. Auch ein Bereich mit niedriger Bandgeschwindigkeit und hoher Heizleistung ist nicht relevant und kann zusätzlich aufgrund der hohen Temperatur und damit steigender Feuergefahr die Anlage gefährden. Um diese Bereiche des Eingangsraums auszuschließen, wurde der zulässige Eingangsraum mittels Randbedingungen in Form zweier Ungleichungen definiert. Die Berücksichtigung dieser Randbedingungen im EDLS-Algorithmus wird in Anhang A.2 erläutert.

Bild 5.5 stellt den inkrementellen Versuchsplan für die drei durchgeführten Vermessungs- und Modellierungsschritte am Bitumenofen dar. Bereiche, die aufgrund der Randbedingungen ausgeschlossen sind, sind grau gekennzeichnet. Im ersten Schritt  $i = 1$  wurde eine Verletzung der zuvor definierten Randbedingungen in Kauf genommen, um den Extrapolationsbereich des ersten Teilmodells möglichst klein zu halten

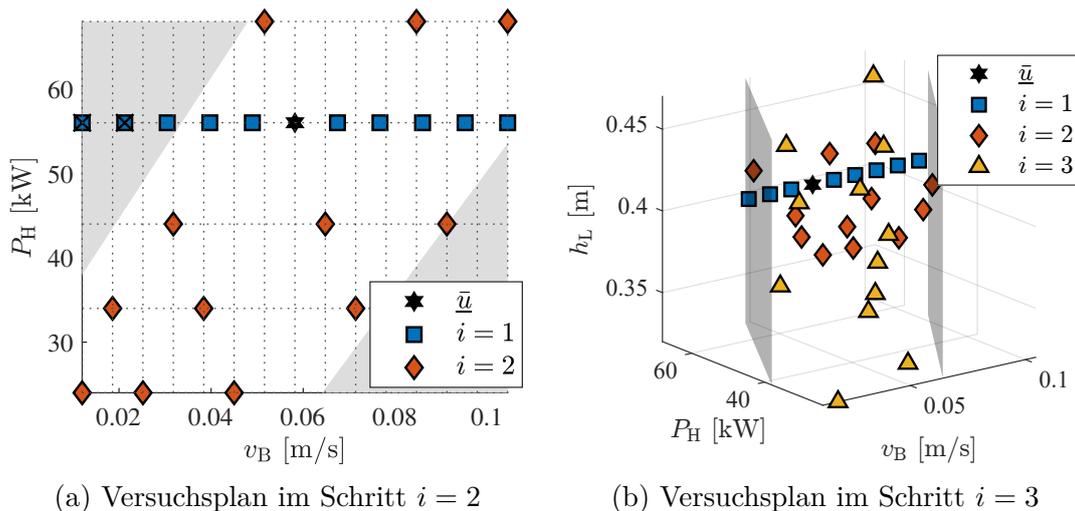


Bild 5.5: Inkrementeller Versuchsplan für die Vermessung des Bitumenofens. Unzulässige Bereiche bzw. die Grenzen des zulässigen Eingangsraums sind grau markiert und nicht vermessene Punkte ausgekreuzt.

(siehe Abschnitt 3.4.3). Aus dem ersten Datenpunkt außerhalb des zuvor definierten Eingangsraums resultierte jedoch bereits eine Überschreitung der zulässigen Temperatur, sodass auf eine Vermessung der weiteren Punkte im Schritt  $i = 1$  mit noch niedrigerer Bandgeschwindigkeit (ausgekreuzte Quadrate) verzichtet wurde.

Im zweiten und dritten Schritt wurden Eingangsgrößen in den Versuchsplan einbezogen, für die durch die Anlagenkonstruktion diskrete Stufen vorgegeben sind. Diese vorgegebenen Stufen müssen folglich auch für den in jedem Schritt erzeugten Latin Hypercube (LHC) verwendet werden. Durch das Konstruktionsprinzip des LHC mit nur einem Punkt pro Stufe wird allerdings die Anzahl der möglichen Versuchspunkte limitiert, sodass mit den fünf diskreten Stufen von  $P_H$  und  $h_L$  nur fünf Versuchspunkte in jedem Schritt realisierbar sind. Um eine größere Anzahl an Versuchspunkten zu ermöglichen, können die vorgegebenen Stufen mehrfach mit Punkten besetzt. Trotz dieser Änderung kann der EDLS-Algorithmus zur Erzeugung eines raumfüllenden Versuchsplans verwendet werden, ohne dass Anpassungen am Algorithmus notwendig sind. Für den Bitumenofen wurden drei Punkte pro Stufe gewählt, wodurch der inkrementelle Versuchsplan über  $\underline{N} = [9 \ 12 \ 12]^T$  Versuchspunkte verfügt.

### Versuchsdurchführung und Modellbildung

Zur Vermessung der einzelnen Punkte gemäß des aufgestellten Versuchsplans wurden die Eingangsgrößen Bandgeschwindigkeit und Heizleistung am Bitumenofen eingestellt und das Abklingen der dynamischen Effekte abgewartet. Falls nötig, wurde durch gezieltes Aufheizen oder Abkühlen eine für den regulären Betrieb typische Ofentemperatur herbeigeführt, sodass das Korrekturmodell gültig ist. Anschließend wurden Laugenbehälter des benötigten Typs einzeln dem Ofen zugeführt und die erzielten Bitumenmattentemperaturen gemessen. Mit zusätzlichen Messungen wurde ein unabhängiger Testdatensatz erstellt. Hierzu wurde ein unter Berücksichtigung der Randbedingungen optimierter LHC mit allen drei betrachteten Eingangsgrößen und  $N_{\text{test}} = 15$  Versuchspunkten verwendet.

Auf Basis des inkrementellen Versuchsplans wurden mehrere inkrementelle Modelle mit Polynommodellen, lokalen Modellnetzen (LMN) und Gaußprozessmodellen (GPM) erstellt. Bild 5.6a zeigt exemplarisch die Trainingsdaten in den Schritten  $i = 1$  und  $i = 2$  und den Modellausgang des GPM-DP Modells für die zwei wichtigsten Eingangsgrößen  $v_B$  und  $P_H$ . Der Modellausgang weist physikalisch sinnvolle Monotonieeigenschaften auf, sodass das Modellverhalten plausibel erscheint. Eine

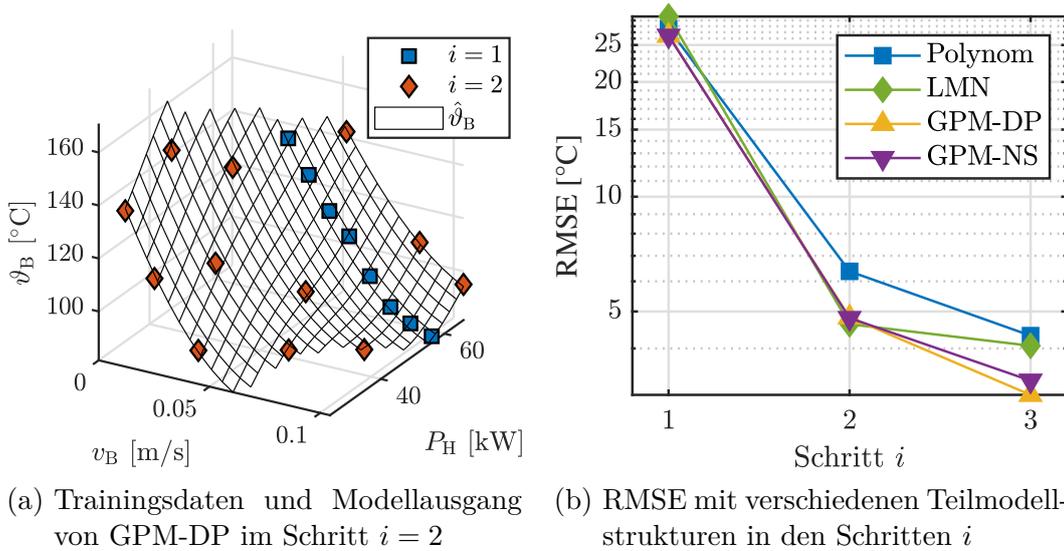


Bild 5.6: Ergebnisse der inkrementellen Modellbildung des Bitumenofens.

genauere Bewertung der Modelle erfolgt auf Grundlage der Testdaten. Die dabei erzielten Fehlerwerte für verschiedene Teilmodellstrukturen sind in Bild 5.6b dargestellt. Bei allen Modellen ist eine Abnahme des RMSE mit einer zunehmenden Anzahl an durchgeführten Schritten sichtbar. Diese Abnahme resultiert aus der Berücksichtigung weiterer Eingangsgrößen, deren Einflüsse durch das Modell abgebildet werden. Alle Teilmodellstrukturen weisen ähnliche Modellfehler auf, wobei das GPM-DP Modell im Schritt  $i = 3$  die besten Ergebnisse erzielt.

## 5.2.2 Implementierung der Optimierungsstrategie

Das erstellte Prozessmodell kann zur Bestimmung optimaler Prozessparameter für den Betrieb des Bitumenofens verwendet werden. Ziel der Optimierung ist es, eine vorgegebene Bitumenmattentemperatur  $\vartheta_{B,\text{soill}}$  für jeden verarbeiteten Laugenbehälter zu erreichen und damit  $|\hat{\vartheta}_B - \vartheta_{B,\text{soill}}|$  zu minimieren. Zusätzlich wird im Sinne eines ökonomischen und ökologischen Anlagenbetriebs ein minimaler Energiebedarf angestrebt. Bei isolierter Betrachtung der Anlage, kann dieses zweite Ziel als eine Minimierung des Energiebedarfs pro verarbeitetem Laugenbehälter

$$E_L = \frac{P_H \cdot t}{D_L(v_B) \cdot t} = \frac{P_H}{D_L(v_B)} \quad (5.2)$$

formuliert werden, wobei  $D_L \propto v_B$  den Durchsatz an Laugenbehältern pro Zeit bezeichnet und vereinfacht die Heizleistung  $P_H$  als Leistung der gesamten Anlage ange-

nommen wird. Für die Lösung des resultierenden multikriteriellen Optimierungsproblems werden die in Bild 5.7 dargestellten Pareto-Optima betrachtet. Für jede der drei häufigsten Laugenbehälterhöhen ist eine Pareto-Front dargestellt, also die Punkte, bei denen keine weitere Verbesserung eines Optimierungskriteriums möglich ist, ohne das andere zu verschlechtern [55]. Am Bitumenofen geht folglich eine Verringerung des Energiebedarfs mit einer größeren Abweichung der Bitumenmattentemperatur einher. Aus diesen Pareto-optimalen Betriebspunkten können anschließend für den Betrieb geeignete Punkte ausgewählt werden. Im Sinne einer möglichst hohen Qualität können Punkte ausgewählt werden, bei denen die gewünschte Bitumenmattentemperatur gemäß Modell mit  $|\hat{\vartheta}_B - \vartheta_{B,\text{soll}}| = 0^\circ\text{C}$  exakt getroffen wird.

### Optimierung unter Berücksichtigung von Produktionsvorgaben

Die zuvor ermittelten Betriebspunkte sind bezogen auf die einzelne Anlage energetisch optimal. Wird der Bitumenofen allerdings in eine Produktionskette eingebunden, müssen weitere Vorgaben bei der Optimierung berücksichtigt werden. Bild 5.8 stellt die Berücksichtigung dieser Produktionsvorgaben am Bitumenofen dar. Diese werden entweder als vorgegebene Größen für das Prozessmodell verwendet oder beeinflussen den Optimierer. Im regulären Betrieb sind die Laugenbehälterhöhe  $h_L$ , die aus dem aktuellen Fertigungsauftrag resultiert, und die in der Anlage gemessene Ofentemperatur  $\vartheta_O$  vorgegebene Größen für das Prozessmodell. Weitere Produktionsvorgaben sind der Sollwert der Bitumenmattentemperatur  $\vartheta_{B,\text{soll}}$  und ein benö-

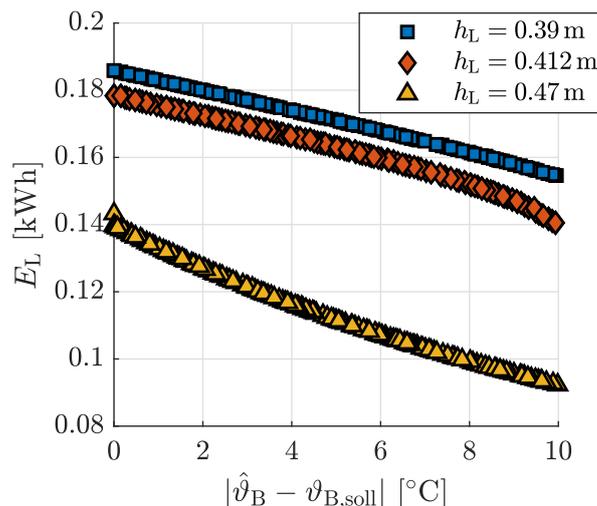


Bild 5.7: Pareto-Optima der multikriteriellen Optimierung des Bitumenofens, bezogen auf einzelne Laugenbehälter. Die Ergebnisse sind für verschiedene Laugenbehälterhöhen  $h_L$  und  $\vartheta_{B,\text{soll}} = 125^\circ\text{C}$  dargestellt.

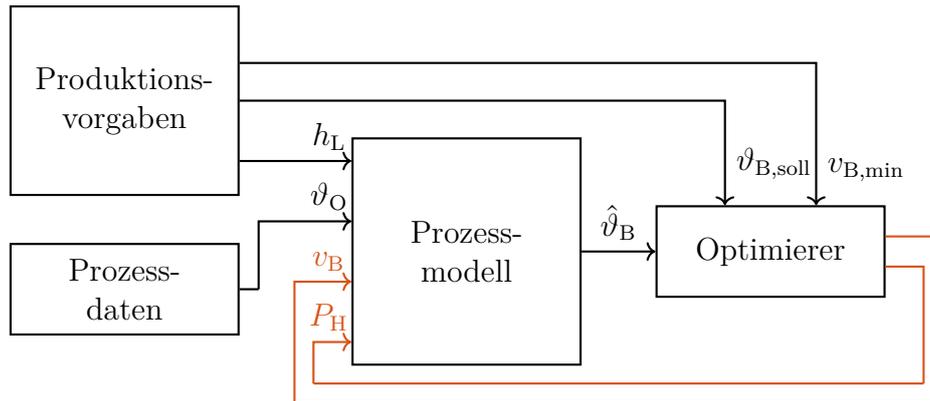


Bild 5.8: Einbeziehung von Produktionsvorgaben bei der Optimierung des Bitumenofens als Eingangsgrößen für das Prozessmodell und als Sollwertvorgabe bzw. Nebenbedingung bei der Optimierung. Die aus der Optimierung resultierenden Eingangsgrößen  $v_B$  und  $P_H$  sind in orange hervorgehoben.

tiger Bauteildurchsatz  $D_L(v_{B,\min})$ , dem eine minimale Bandgeschwindigkeit  $v_{B,\min}$  zugeordnet werden kann. Die Bandgeschwindigkeit muss mindestens  $v_{B,\min}$  entsprechen, damit der Bitumenofen keinen Flaschenhals in der Produktionskette darstellt. Eine Steigerung der Bandgeschwindigkeit über den Minimalwert hinaus führt hingegen zu keinem gesteigerten Durchsatz  $D_L(v_B) = D_L(v_{B,\min})$  für  $v_B > v_{B,\min}$ , da der Gesamtdurchsatz in einer verketteten Produktion durch die Anlage mit dem geringsten Durchsatz bestimmt wird. Das Kriterium eines minimalen Energiebedarfs pro verarbeitetem Laugenbehälter gemäß Gleichung (5.2) kann bei einem solchen konstanten Durchsatz als ein Betrieb mit kleinstmöglicher Leistung  $P_H$  formuliert werden. Ziel des Optimierers ist die Bestimmung der kleinstmöglichen Heizleistung  $P_H$  und der dazugehörigen Bandgeschwindigkeit  $v_B$ , die gemäß des Prozessmodells zu der vorgegebenen Bitumenmattentemperatur  $\hat{\vartheta}_B = \vartheta_{B,\text{soll}}$  führen.

Die realisierte Optimierungsstrategie ist in Bild 5.9 gezeigt. Dargestellt sind alle Kombinationen der Eingangsgrößen  $P_H$  und  $v_B$ , die gemäß des Prozessmodells bei festen Produktionsvorgaben zu  $\hat{\vartheta}_B = \vartheta_{B,\text{soll}}$  führen. Um die minimale Bandgeschwindigkeit  $v_{B,\min}$  und die diskreten Heizstufen zu berücksichtigen, werden nacheinander folgende Schritte durchgeführt:

- i Bestimmung der minimalen Heizleistung bei gegebener minimaler Bandgeschwindigkeit  $P_{H,\min} = \arg \min_{P_H} (|\hat{\vartheta}_B(v_{B,\min}, P_H, h_L, \vartheta_O) - \vartheta_{\text{soll}}|)$ .
- ii Wahl der Heizstufe  $P_H = \min(P_H \in \{24, 34, 44, 56, 68\} \text{ kW} | P_H \geq P_{H,\min})$  basierend auf der zuvor ermittelten minimalen Heizleistung.

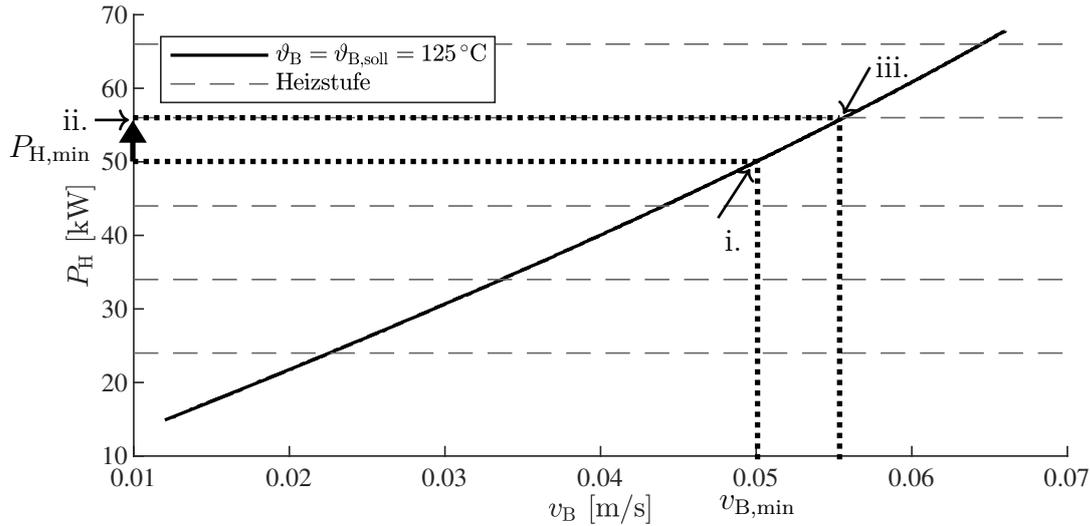


Bild 5.9: Durch Optimierung am Modell ermittelte Kombinationen der Eingangsgrößen  $v_B$  und  $P_H$  für die  $\vartheta_B = \vartheta_{B,soll} = 125^\circ\text{C}$  gilt. Exemplarisch wurden die Produktionsvorgaben  $h_L = 0.412\text{ m}$ ,  $\vartheta_O = 80^\circ\text{C}$  und  $v_{B,min} = 0.05\text{ m/s}$  zugrunde gelegt. Die Optimierungsstrategie besteht aus drei Schritten (i.-iii.).

- iii Ermittlung der optimalen Bandgeschwindigkeit für die gewählte Heizstufe  $v_B = \arg \min_{v_B} (|\hat{\vartheta}_B(v_B, P_H, h_L, \vartheta_O) - \vartheta_{soll}|)$ .

Es ergeben sich ausschließlich einkriterielle Optimierungsprobleme in den Schritten i. und iii., die mittels Newton-Verfahren [92] gelöst werden.

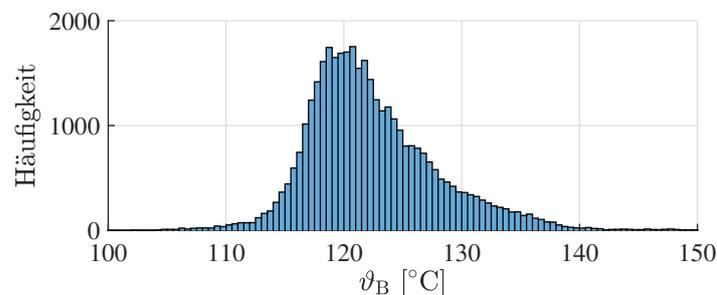
### Implementierung und Evaluation im laufenden Prozessbetrieb

Um den regulären Anlagenbetrieb zu verbessern, wurde die zuvor beschriebene Optimierungsstrategie in der *speicherprogrammierbaren Steuerung* (SPS) der Anlage implementiert. Eine solche Implementierung kann auf verschiedene Weise erfolgen, wobei Beschränkungen der Anlagensteuerung hinsichtlich Rechenleistung und Speicherkapazität zu berücksichtigen sind. Ein übliches Vorgehen, um rechenaufwändige Optimierungsstrategien trotz dieser Beschränkungen zu realisieren, ist das Abbilden von offline ermittelten Optimierungsergebnisse mit einem datenbasierten Modell (*Optimalsteuerung*). In [96] werden für ein solches Vorgehen Kennfelder, lokale Modellnetze und Radial-Basisfunktionen-Netze für die Implementierung in Motorsteuergeräten verglichen. Für den Bitumenofen wurde aufgrund des akzeptablen Rechenaufwands der entwickelten Optimierungsstrategie eine Online-Optimierung gewählt, wozu das inkrementelle Polynommodell des Bitumenofens und die Optimierungsstrategie in einen Programmbaustein für die SPS überführt wurden. Das

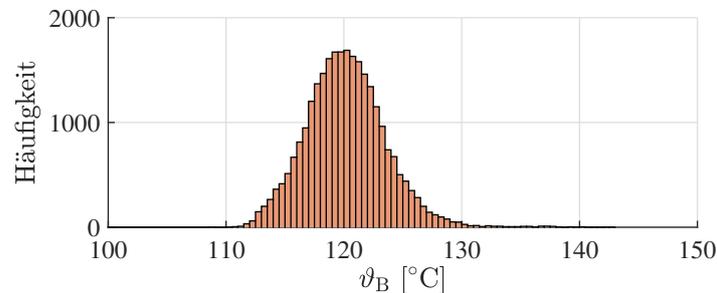
Polynommodell wurde hierbei aufgrund der einfachen Implementierung bei zugleich hinreichender Modellgüte ausgewählt.

Bild 5.10 stellt die gemessene Bitumenmattentemperatur im regulären Prozessbetrieb über einen Zeitraum von jeweils 5 Wochen vor und nach der Inbetriebnahme des Optimierungsbausteins gegenüber. Hierbei werden nur diejenigen Bauteile betrachtet, bei denen für mindestens vier Minuten eine konstante Heizleistung vorliegt, um dynamische Effekte durch das Aufheizen oder Abkühlen der Heizstrahler auszuschließen. Im Vergleich der Histogramme ist im Betrieb mit aktiviertem Optimierungsbaustein eine geringere Streuung der Bitumenmattentemperatur sichtbar. Insbesondere hohe Temperaturen mit  $\vartheta_B > 130^\circ\text{C}$  treten vermindert auf. Diese Reduktion von hohen Temperaturen führt zugleich zu einer Verringerung des Energiebedarfs. So wird der mittlere Energiebedarf pro Laugenbehälter in dem betrachteten Zeitraum durch den Optimierungsbaustein um ca. 8 % reduziert.

Darüber hinaus erhöht der implementierte Optimierungsbaustein die Flexibilität der Anlage. So können auch andere Typen von Laugenbehältern mit einer Höhe im vermessenen Wertebereich verarbeitet werden, ohne dass neue Messungen oder Einstellarbeiten notwendig sind. Außerdem werden Einschränkungen im Anlagenbetrieb wie



(a) Ohne Optimierungsbaustein: Bitumenmattentemperaturen von ca. 37000 Bauteilen



(b) Mit Optimierungsbaustein: Bitumenmattentemperaturen von ca. 27000 Bauteilen

Bild 5.10: Histogramme der Bitumenmattentemperatur im regulären Prozessbetrieb ohne und mit Optimierungsbaustein.

ein Ausfall von Heizstrahlern, der zu veränderten Heizstufen führt, bei der Optimierung berücksichtigt und automatisch kompensiert.

### 5.3 Qualitätsvorhersage für die Montage von Waschmaschinentrommeln

Als industrielle Anwendung für die Versuchsplanung für zusammengesetzte Einheiten wird in diesem Abschnitt ein Montageprozess von Waschmaschinentrommeln der Firma Miele & Cie. KG betrachtet. Waschmaschinentrommeln sind Baugruppen, die aus mehreren Komponenten zusammengesetzt werden und an die hohe Qualitätsanforderungen gestellt werden. So sind insbesondere Rund- und Planlauf der Trommel wichtige Qualitätsgrößen. Zu hohe Abweichungen können dazu führen, dass Wäsche während des Waschvorgangs unter die Trommel gezogen wird.

Bild 5.11 stellt die wichtigsten Komponenten einer Waschmaschinentrommel zusammen mit den nachfolgend betrachteten Qualitätsgrößen schematisch dar. Bei den Hauptkomponenten einer Trommel handelt es sich um Kappe ( $c_1$ ), Mantel ( $c_2$ ), Boden ( $c_3$ ) und Flansch ( $c_4$ ). Diese Komponenten sind radialsymmetrisch und können

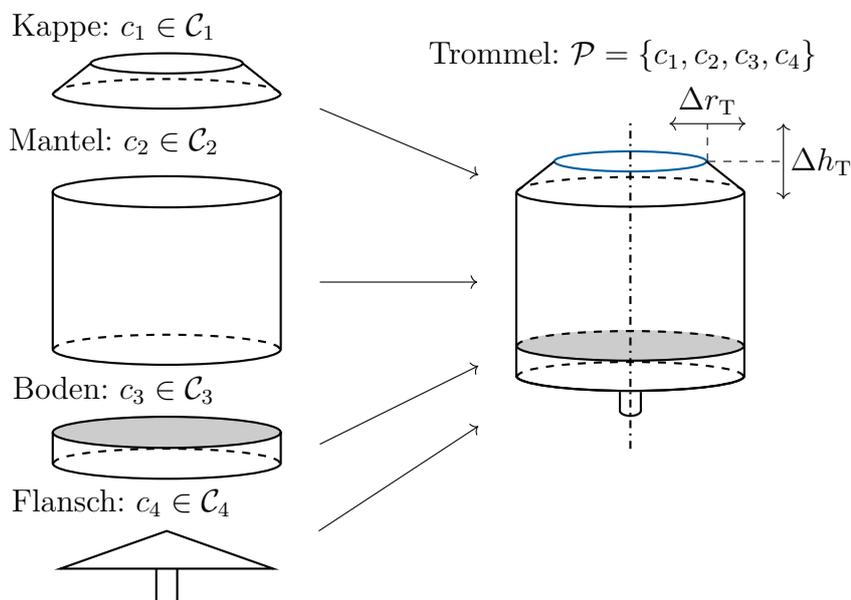


Bild 5.11: Schematische Darstellung der Hauptkomponenten einer Waschmaschinentrommel. Die Qualitätsgrößen beschreiben die Spanneiten der Abweichungen über den Umfang der Trommelöffnung (blau) in radialer  $\Delta r_T$  und axialer  $\Delta h_T$  Richtung.

daher in unterschiedlichen rotatorischen Ausrichtungen verbaut werden. Die Montage erfolgt in einem vollautomatisierten Prozess, der die Komponenten fest miteinander verbindet. Eine montierte Trommel  $\mathcal{P} = \{c_1, c_2, c_3, c_4\}$  besitzt eine Rotationsachse (Strichpunktlinie). Die Rundlaufabweichung  $\Delta r_T$  beschreibt die Spannweite der Abweichungen der Kappenöffnung (blau) in radialer Richtung, während die Planlaufabweichung  $\Delta h_T$  die Spannweite der axialen Abweichungen der Kappenöffnung angibt.

Mittels eines Modells sollen die Rund- und Planlaufabweichungen der montierten Waschmaschinentrommeln aus Merkmalen der verwendeten Komponenten vorhergesagt werden. Hierzu erfolgt eine experimentelle Modellbildung auf Basis der in Kapitel 4 beschriebenen Versuchsplanungsmethodik. Dieses Vorgehen wurde gegenüber einer Modellierung auf Basis von Prozessdaten aus dem regulären Anlagenbetrieb gewählt, da es in dem betrachteten Anwendungsfall folgende Vorteile aufweist:

- Steigerung der Eingangsraumabdeckung durch eine gezielte Zuordnung der Einzelkomponenten und eine Vorgabe der rotatorischen Ausrichtung.
- Geringere Anforderungen an die Datenerfassung durch manuelle Bauteilzuführung (z. B. keine Notwendigkeit der Rückverfolgbarkeit der Komponenten).
- Frühzeitige Modellerstellung bereits in der Inbetriebnahmephase eines Montageprozesses.

Ein Modell zur Vorhersage der Abweichungen einer Trommel kann direkt genutzt werden, um diese Abweichungen zu reduzieren. Dies ermöglicht es, die Qualität der Baugruppen zu steigern sowie Ausschuss bzw. Nacharbeit zu reduzieren. Hierzu können verschiedene Kombinationen von Komponenten bzw. Chargen an Komponenten am Modell evaluiert werden. Die Kombination mit der geringsten vorhergesagten Abweichung kann anschließend montiert werden. Eine solche gezielte Zusammenstellung der Komponenten würde allerdings den logistischen Aufwand in der Produktion erhöhen. Eine weniger aufwändige Möglichkeit zur Steigerung der Qualität ist eine gezielte rotatorische Ausrichtung der Komponenten zueinander. Typischerweise ist in der Praxis eine geringe Anzahl diskreter Winkelpositionen vorteilhaft, für die mittels vollständiger Enumeration optimale Rotationswinkel am Modell ermittelt werden können.

Um ein solches Modell zu erstellen, werden nachfolgend zwei verschiedene Ansätze betrachtet, die schwerpunktmäßig auf die Vorhersage einer der Qualitätsgrößen abzielen und auf konstruktiv unterschiedliche Typen an Waschmaschinentrommeln

angewandt wurden. Der in Abschnitt 5.3.1 beschriebene lokale Ansatz wird zur Vorhersage der Trommelhöhe an diskreten Positionen entlang der Trommelöffnung verwendet und charakterisiert die Planlaufabweichung. Durch die Vernachlässigung von Interaktionen zwischen den einzelnen Positionen wird eine geringe Dimensionalität und ein hoher Trainingsdatenumfang erzielt. Der in Abschnitt 5.3.2 beschriebene globale Ansatz ist hingegen in der Lage, Interaktionen zwischen den einzelnen Messpositionen einzubeziehen. Hierfür werden globale Merkmale der Einzelkomponenten verwendet und für die Vorhersage der Rundlaufabweichung genutzt. Erste Ergebnisse zu dieser Anwendung wurden in den Konferenzbeiträgen [115, 116] veröffentlicht.

### 5.3.1 Lokaler Ansatz zur Vorhersage von Höhenabweichungen

In diesem Abschnitt wird der lokale Ansatz gemäß Abschnitt 4.5.2 zur Vorhersage von Höhenabweichungen einer montierten Trommelbaugruppe verwendet. Diesem Ansatz liegt die Annahme zugrunde, dass sich die Trommelhöhe an einer Winkelposition hauptsächlich aus den Merkmalen der verwendeten Komponenten an dieser Winkelposition ergibt. Ist beispielsweise die Höhe einer Kappe an einer Winkelposition vergrößert, wird angenommen, dass dies zu einer Vergrößerung der Trommelhöhe an dieser Winkelposition beiträgt. Auswirkungen auf die Trommelhöhe an anderen Winkelpositionen werden aufgrund der Verformbarkeit der Komponenten im Montageprozess vernachlässigt.

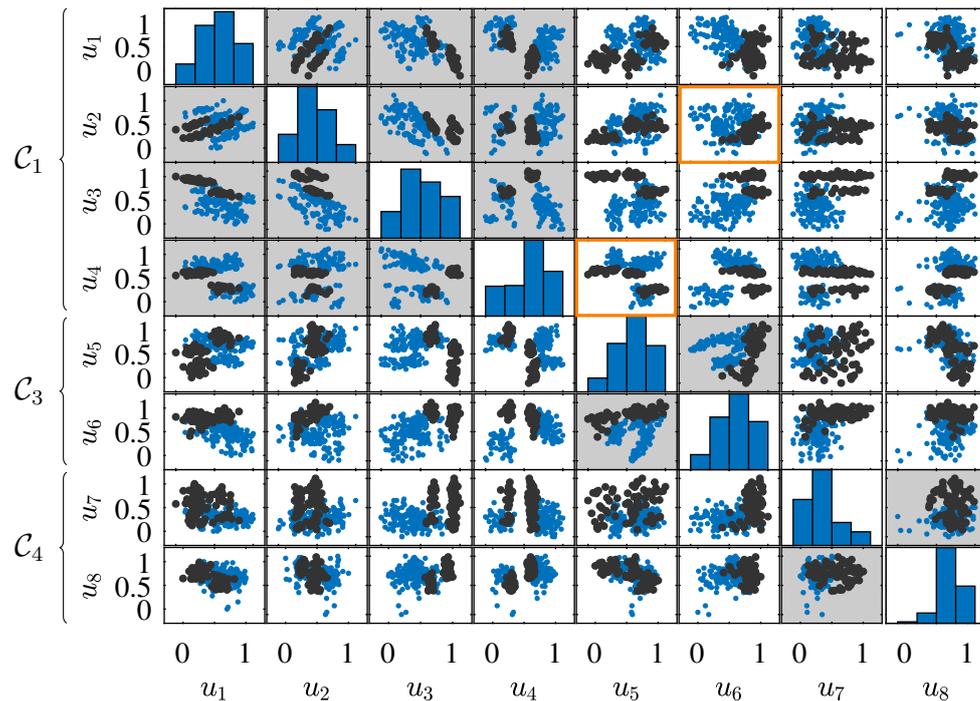
#### Versuchsplanung

Der Versuchsplan basiert auf einem historischen Datensatz mit 25 zufällig hergestellten Trommeln ohne Rotation der Komponenten. Dieser gegebene Datensatz wird um zusätzliche 60 Trommeln erweitert, wobei sowohl die Abstände der neuen Versuchspunkte untereinander als auch die Abstände zu den gegebenen Datenpunkten optimiert werden. Für alle Komponenten werden  $w = 3$  Winkelpositionen  $0^\circ$ ,  $120^\circ$  und  $-120^\circ$  betrachtet, in denen diese verbaut werden können. An diesen Winkelpositionen werden jeweils 8 verschiedene Merkmale gemessen, die beispielsweise Höhen oder Außenradien von Komponenten beschreiben. Somit ergeben sich insgesamt  $M = 24$  Merkmale der Komponenten, die bei einem globalen Modell zu einer hohen Dimensionalität führen würden. Der lokale Ansatz betrachtet nur die  $\frac{M}{w} = 8$  lokalen Merkmale und führt zugleich bei  $N = 85$  Trommeln zu einer Steigerung der für die Modellschätzung verfügbaren Datenpunkte auf  $N \cdot w = 255$ . Von den lokalen

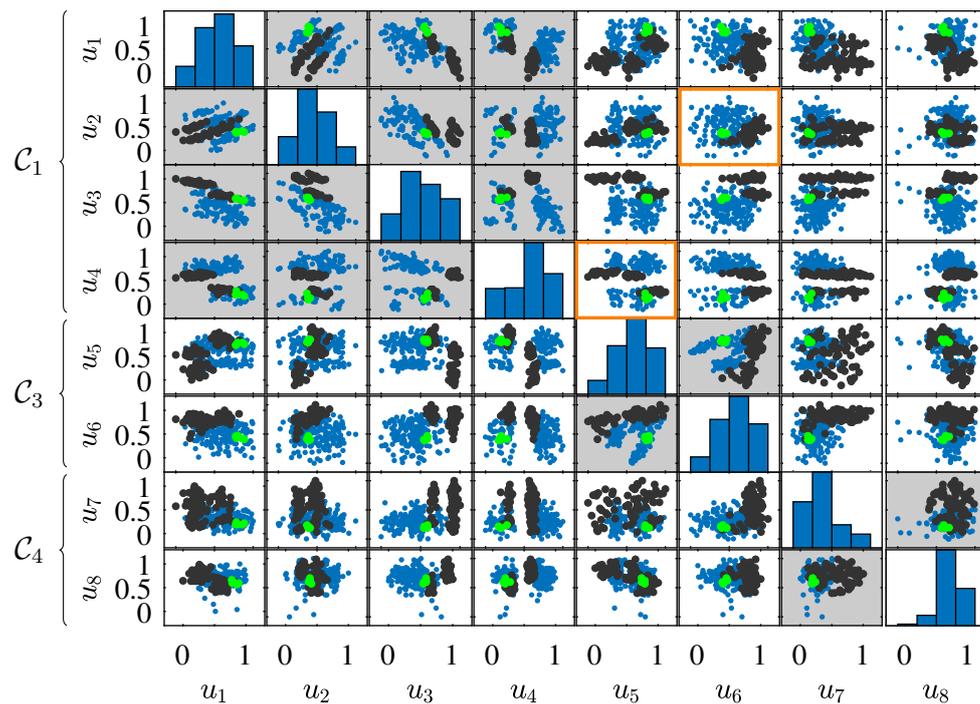
Merkmale entfallen  $m_1 = 4$  auf die Kappe,  $m_3 = 2$  auf den Boden und  $m_4 = 2$  auf den Flansch, wobei ausschließlich geometrische Maße der Komponenten betrachtet werden. Der Mantel bleibt mit  $m_2 = 0$  Merkmalen unberücksichtigt, da von diesem Komponententyp die geringsten Auswirkungen auf die Trommel erwartet werden.

Bild 5.12 stellt in (a) eine initiale Datenverteilung von zufällig zusammengestellten Baugruppen ohne Rotation der Komponenten dar, während in (b) der gemäß dem lokalen Ansatz optimierte Versuchsplan unter Berücksichtigung der Rotation von Komponenten zu sehen ist. Hierbei sind Datenpunkte, die aus dem historischen Datensatz stammen, schwarz markiert. Diese Datenpunkte werden als nicht veränderbare Punkte in alle Versuchspläne übernommen und bei der Optimierung der neuen Versuchspunkte (blau) berücksichtigt. Alle grau hinterlegten Felder können durch die Optimierung nicht verändert werden, da diese ausschließlich Merkmale eines Komponententyps zeigen. So beschreiben zum Beispiel die Merkmale  $u_1, u_2, u_3$  und  $u_4$  unterschiedliche Bauteilmaße des Komponententyps Kappe ( $\mathcal{C}_1$ ). Aufgrund der Kopplung dieser Merkmale an das jeweilige physikalische Objekt sind keine Tauschoperationen möglich, die eine Untermenge dieser Merkmale betreffen. In allen anderen Feldern ist durch die Optimierung eine verbesserte Raumabdeckung sichtbar. Die gleichmäßigere Verteilung der Punkte in den dargestellten Projektionen wird besonders in den orange markierten Feldern deutlich. Zusätzlich wurden im optimierten Versuchsplan  $K = 5$  möglichst ähnliche Datenpunkte erzeugt (siehe Abschnitt 4.3), die in grün dargestellt sind. Um keine Details der Bauteilvermessung der Firma Miele & Cie. KG preiszugeben, sind alle betrachteten Eingangsgrößen auf den Wertebereich  $[0, 1]$  normiert dargestellt.

Zur Quantifizierung der Raumabdeckung werden in Tabelle 5.2 verschiedene Versuchspläne für die realen Komponenten anhand der in Abschnitt 2.3 beschriebenen Bewertungsmaße verglichen, wobei bei beiden Bewertungsmaßen kleinere Werte bessere Ergebnisse charakterisieren. Hierbei werden 1) der zufällige Versuchsplan ohne Rotation von Komponenten aus Bild 5.12a sowie 2) ein weiterer zufälliger Versuchsplan, bei dem zusätzlich eine zufällige Rotation der Komponenten erfolgt, einem 3) mittels lokaler Suche optimierten Versuchsplan gegenübergestellt. Zusätzlich ist 4) der in den Versuchen verwendete Versuchsplan aus Bild 5.12b mit 5 möglichst ähnlichen Versuchspunkten aufgeführt. Es zeigt sich, dass die Rotation von Komponenten durch die zusätzlichen Freiheiten bei der Versuchsplanung zu einer deutlichen Verbesserung der Versuchspläne gemäß des Kriteriums  $D_{\text{KL}}$  führt. Die besten Ergebnisse mit dem geringsten Wert von  $D_{\text{KL}}$  werden durch 3) den lokalen Suchalgorithmus erzielt. Bei 4) führt die Erzeugung ähnlicher Versuchspunkte durch die



(a) Zufälliger Versuchsplan ohne Rotation



(b) Optimierter Versuchsplan mit Rotation

Bild 5.12: Gegenüberstellung von Datenverteilungen gemäß dem lokalen Ansatz (a) ohne und (b) mit Optimierung. Der historische Datensatz (schwarz) wurde um raumfüllende (blau) und möglichst ähnliche Punkte (grün) ergänzt. Nicht optimierbare Felder sind grau hinterlegt, während der Nutzen der Optimierung in den orange markierten Feldern besonders deutlich wird.

Tabelle 5.2: Bewertung von Versuchsplänen gemäß des lokalen Ansatzes mittels  $D_{\text{KL}}$  für die Gleichförmigkeit (siehe Abschnitt 2.3.3) und  $d_{\text{S}}$  für die größten Löcher im Eingangsraum (siehe Abschnitt 2.3.2). Die Ergebnisse der zufälligen Versuchspläne sind über 100 Wiederholungen gemittelt.

Nr.	Zuordnung der Komponenten	Rotation	ähnliche Punkte	$D_{\text{KL}}$	$d_{\text{S}}$
1)	zufällig	keine	keine	8.30	1.27
2)	zufällig		keine	7.08	1.24
3)	lokale Suche		keine	<b>7.06</b>	1.24
4)	lokale Suche		5	7.11	<b>1.15</b>

hiermit verbundenen Einschränkungen zu einer leichten Verschlechterung dieses Kriteriums. Es ist allerdings zu beachten, dass durch die Dimensionalität von acht ein großer Fehler bei der Kerndichteschätzung gegeben sein kann [95], die dem Kriterium  $D_{\text{KL}}$  zugrunde liegt. Insgesamt verdeutlichen der von null verschiedene Wert von  $D_{\text{KL}}$  sowie die Darstellung in Bild 5.12b, dass die erzielten Datenverteilungen deutlich von einer Gleichverteilung abweichen. Ursächlich hierfür sind Einschränkungen durch die gegebenen Komponenten. Es liegt allerdings eine Punktverteilung vor, die unter Berücksichtigung der gegebenen Komponenten möglichst große Bereich des Eingangsraums möglichst gleichmäßig abdeckt.

Bezogen auf die größte Lücke im Eingangsraum sind nur geringe Veränderungen des Kriteriums  $d_{\text{S}}$  zwischen den zufälligen Versuchsplänen und dem lokalen Suchalgorithmus sichtbar. Eine deutliche Verbesserung wird in dem betrachteten Fall bei der lokalen Suche mit zusätzlicher Erzeugung von ähnlichen Punkten erzielt. Im Allgemeinen ist jedoch mit dem gegenteiligen Effekt zu rechnen, da die Erzeugung ähnlicher Datenpunkte zu Einschränkungen in der Versuchsplanung führt, die Lücken im Eingangsraum tendenziell vergrößern.

### Versuchsdurchführung und Modellbildung

In den Versuchen wurden die Komponenten in den durch den Versuchsplan vorgegebenen Kombinationen und rotatorischen Ausrichtungen der Montageanlage zugeführt. Innerhalb der Anlage erfolgte die vollautomatische Montage zu Waschmaschinentrommeln mit einer anschließenden Messung der Abweichungen dieser Trommeln. Um eine Zuordnung der Messwerte zu den einzelnen Baugruppen zu ermöglichen, wurden diese manuell gekennzeichnet.

Auf Basis der Versuchsdaten wird zur Vorhersage jeder Qualitätsgröße ein separates Gaußprozessmodell (GPM) erstellt. Die acht Eingangsgrößen für diese Modelle sind durch den Versuchsplan vorgegeben, während die jeweils gemessenen Abweichungen der Trommelhöhe bzw. die radiale Abweichung der Trommelöffnung als Ausgangsgrößen verwendet werden. Die Validierung der Modelle erfolgt mittels *Leave-One-Out-Kreuzvalidierung*, bei der für einen Validierungsdatenpunkt jeweils eine Vorhersage durch ein Modell erfolgt, das auf den verbleibenden  $N - 1$  Datenpunkten trainiert wurde. Dieser Vorgang wird  $N$  mal wiederholt, wobei jeder Datenpunkt einmal die Rolle des Validierungsdatenpunkts einnimmt [80]. Der RMSE über alle Validierungsdatenpunkte wird als Leave-One-Out-Fehler bezeichnet.

Die erzielten Ergebnisse sind in Tabelle 5.3 zusammengefasst. Bei der radialen Abweichung  $r_T$  zeigt sich anhand der nahezu identischen Werte von Leave-One-Out-Fehler und Standardabweichung, dass das erstellte Modell eine vergleichbare Modellgüte wie ein Mittelwertmodell der Ausgangsgröße aufweist. Da bei einem solchen Mittelwertmodell die Eingangsgrößen unberücksichtigt bleiben, kann geschlussfolgert werden, dass diese Größe mit dem lokalen Ansatz und den ausgewählten Merkmalen nicht vorhergesagt werden kann. Die Vernachlässigung der Interaktionen zwischen den einzelnen Positionen beim lokalen Ansatz kann hierfür ursächlich sein.

Für die Trommelhöhe  $h_T$  ergibt sich ein um 33% reduzierter Fehler im Vergleich zur Standardabweichung, was darauf schließen lässt, dass das Modell funktionale Zusammenhänge des Prozesses erlernt hat. In der Anwendung des Modells soll eine optimale Ausrichtung gegebener Komponenten bestimmt werden, bei der Abweichungen der Trommelhöhe minimal sind. Hierbei könnte ein praktischer Einsatz des Modells trotz des großen Fehlers zielführend sein, da nur die korrekte Ausrichtung, aber nicht der Wert des vorhergesagten Optimums von Bedeutung ist. Eine Validierung des Modells im regulären Betrieb konnte allerdings aufgrund der fehlenden Rückverfolgbarkeit der Bauteile an der betrachteten Anlage nicht umgesetzt werden.

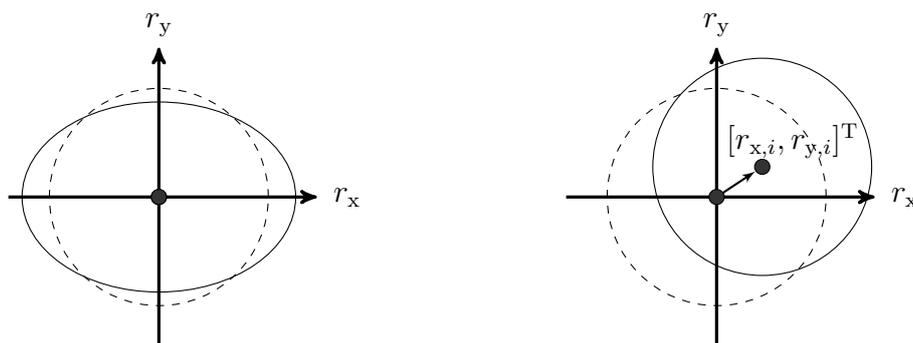
Tabelle 5.3: Vergleich des Leave-One-Out-Fehlers mit der Standardabweichung der beiden betrachteten Qualitätsgrößen.

Maß	Radiale Abweichung $r_T$	Trommelhöhe $h_T$
Leave-One-Out-Fehler	0.254	0.163
Standardabweichung	0.253	0.240

### 5.3.2 Globaler Ansatz zur Vorhersage von Rundlaufabweichungen

In diesem Abschnitt wird der globale Ansatz zur rotatorischen Ausrichtung von Komponenten (vgl. Abschnitt 4.5.1) verwendet. Hierbei werden die mittleren Abweichungen der einzelnen Komponenten durch globale Merkmale charakterisiert. Als Anwendung dieses Ansatzes wird ein Montageprozess für einen neu entwickelten Typ von Waschmaschinentrommeln betrachtet. Dieser Montageprozess dient zur Erstellung einer Teilbaugruppe ohne montierten Flansch, sodass nur die Komponenten Kappe, Mantel und Boden montiert werden. Durch ein geändertes Montageverfahren ist davon auszugehen, dass Abweichungen der Trommelhöhe im Vergleich zu dem in Abschnitt 5.3.1 betrachteten Montageprozess eine geringe Relevanz haben. Somit können die Rundlaufabweichungen als entscheidende Größe für die Gesamtqualität der Trommeln betrachtet werden. Diese Art von Abweichung steht daher in der nachfolgend beschriebenen experimentellen Modellbildung im Fokus.

Bild 5.13 stellt zwei verschiedene Ursachen für das Auftreten von Rundlaufabweichungen dar, die durch die verwendeten Komponenten hervorgerufen werden können. Zusätzlich können Überlagerungen dieser beiden Ursachen auftreten. In (a) ist schematisch die Kontur einer Komponente dargestellt, deren Mittelpunkt mit der Konstruktionsvorgabe übereinstimmt, die aber von der vorgegebenen Kreisform abweicht. Hieraus resultiert ein Einfluss der Komponente auf die Baugruppe, der sich an den einzelnen Winkelpositionen unterschiedlich auswirkt. Im dargestellten Beispiel ist dieser Einfluss z. B. in Richtung der  $r_x$ -Achse positiv (größere Ausdehnung als vorgesehen) und in Richtung der  $r_y$ -Achse negativ (kleinere Ausdehnung als vor-



(a) Rundlaufabweichungen durch unrunde Komponenten      (b) Rundlaufabweichungen als Resultat von Konzentritätsabweichungen

Bild 5.13: Darstellung zweier Ursachen für Rundlaufabweichungen. Die Kontur einer Komponente (durchgezogene Linie) weicht auf verschiedene Arten von der Konstruktionsvorgabe (gestrichelte Linie) ab.

gesehen). Für die Abbildung eines solchen lokal unterschiedlichen Einflusses ist der in Abschnitt 5.3.1 beschriebene lokale Ansatz geeignet. Allerdings zeigen die Ergebnisse in Tabelle 5.3, dass bei einem vergleichbaren Trommeltyp mit dem lokalen Ansatz keine sinnvollen Zusammenhänge abgebildet werden können.

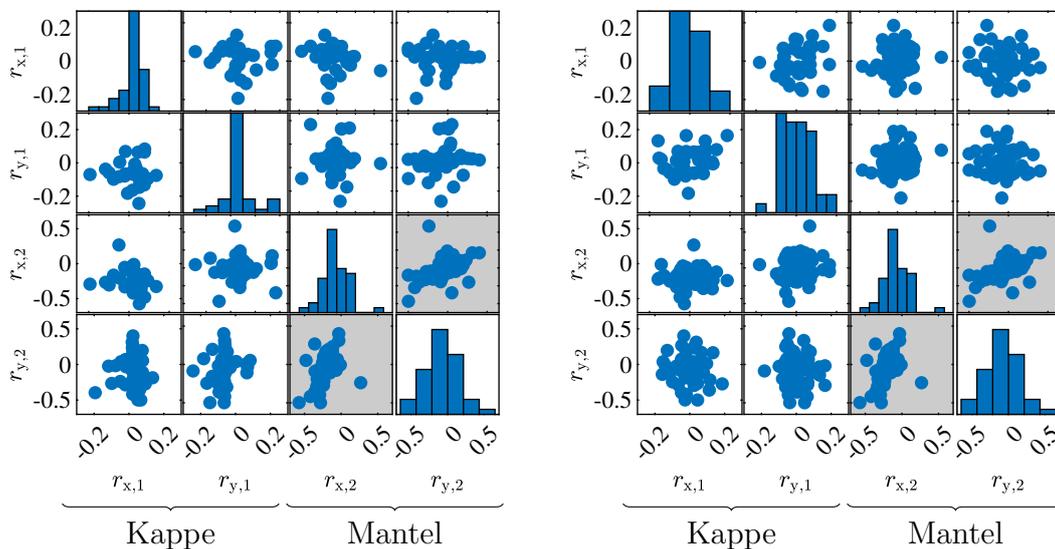
Als Hauptursache für Rundlaufabweichungen kann die in (b) dargestellte Konzentritätsabweichung von Komponenten angenommen werden, die einen Versatz der Kreismittelpunkte beschreibt. Da eine Trommel aus mehreren Komponenten besteht, die aufeinander montiert werden, kann jede dieser Komponenten mit ihrer Konzentritätsabweichung einen radialen Versatz hervorrufen, der sich auf die Baugruppe auswirkt. Um diese Einflüsse zu erfassen und zu modellieren, wird der globale Ansatz aus Abschnitt 4.5.1 verwendet. Hierbei beschreibt jeweils ein Merkmalsvektor  $[r_{x,i} \ r_{y,i}]^T$  die Abweichung des Kreismittelpunkts vom Nennmittelpunkt jeder einzelnen Komponente des Typs  $i$ . Der Kreismittelpunkt errechnet sich als Ausgleichskreis aus mehreren Messpunkten (siehe Bild 4.16). Je nach Ausrichtung, in der eine Komponente in der Baugruppe verbaut wird, wirkt sich diese radiale Abweichung unterschiedlich auf die Baugruppe aus. Ein Ausgangsgrößenvektor  $[r_{x,T} \ r_{y,T}]^T$  ergibt sich analog zur Bestimmung der Merkmalsvektoren aus den entlang der Trommelöffnung gemessenen radialen Abweichung  $r_T$ . Dieser Vektor charakterisiert die radiale Abweichung der gesamten Trommelbaugruppe.

## Versuchsplanung

Für die Versuche standen insgesamt 45 Komponenten jedes Typs zur Verfügung. Aus diesen Komponenten konnten maximal 45 Baugruppen montiert werden, die mit dem globalen Ansatz zu  $N = 45$  Datenpunkten führen. Der Komponententyp Boden bleibt zugunsten einer niedrigen Dimensionalität des Eingangsraums im Versuchsplan unberücksichtigt. Diese Vernachlässigung wird aufgrund einer hohen Reproduzierbarkeit bei der Herstellung der Böden und den geringen radialen Abweichungen dieses Komponententyps gewählt. In den Versuchen werden alle Böden in derselben rotatorischen Ausrichtung in den Baugruppen verbaut. Da jede der zwei betrachteten Komponenten Kappe und Boden durch einen zweidimensionalen Merkmalsvektor beschrieben wird, ergibt sich ein Eingangsraum der Dimensionalität  $M = 4$ . Um die Streuung der Merkmalsausprägungen bei den Kappen zu vergrößern, wurde die Hälfte der verfügbaren Kappen vor deren Vermessung manuell in zufälliger radialer Richtung verformt. Durch die Versuchspläne werden sowohl gezielte Zuordnungen

der Kappen ( $\mathcal{C}_1$ ) und Mäntel ( $\mathcal{C}_2$ ) beschrieben als auch die rotatorischen Ausrichtungen der Kappen relativ zu den Mänteln vorgegeben. Aus praktischen Gründen werden nur diskrete Winkelpositionen in  $60^\circ$ -Schritten zugelassen. Bild 5.14 zeigt die resultierende Datenverteilung in (a) vor und in (b) nach der Optimierung mit dem lokalen Suchalgorithmus unter Einbeziehung der Rotation von Komponenten. Durch die Optimierung wird in (b) eine gleichmäßigere Datenverteilung erzielt. Allerdings sind weiterhin die Bereiche des Eingangsraums, die mit Versuchspunkten abgedeckt werden können, durch die verfügbaren Komponenten eingeschränkt.

Auf Basis der Komponenten werden verschiedene Versuchspläne verglichen, die in Tabelle 5.4 anhand der in Abschnitt 2.3 beschriebenen Bewertungsmaße gegenübergestellt sind. Hierbei werden zwei Versuchspläne mit zufälliger Zuordnung der Komponenten zu den Baugruppen betrachtet, wobei in 1) keine Rotation der Komponenten erfolgt, während in 2) eine zufällige rotatorische Ausrichtung der Komponenten gegeben ist. Durch die Einbeziehung der Rotation wird die Gleichförmigkeit der Datenverteilung gemäß des Kriteriums  $D_{\text{KL}}$  verbessert und zusätzlich die größte Lücke im Eingangsraum gemäß  $d_S$  verkleinert. Eine weitere Verbesserung von  $D_{\text{KL}}$  ist in 3) durch die Anwendung der lokalen Suche gegeben, wobei  $d_S$  einen nahezu identischen Wert wie in 2) aufweist. Auf den Versuchsplan in 3) wird die in Abschnitt 4.4



(a) Zufällige Zuordnung der Komponenten ohne Rotation (b) Lokale Suche mit Rotation der Komponenten

Bild 5.14: Datenverteilung gemäß des globalen Ansatzes vor und nach der Optimierung. Felder mit ausschließlichen Merkmalen des Mantels sind unveränderbar (grau), da der Mantel als Bezug für die Rotation dient. Bei der Kappe können alle Felder durch die Rotation verändert werden.

Tabelle 5.4: Bewertung von Versuchsplänen gemäß des globalen Ansatzes mittels  $D_{\text{KL}}$  für die Gleichförmigkeit (siehe Abschnitt 2.3.3) und  $d_S$  für die größten Löcher im Eingangsraum (siehe Abschnitt 2.3.2). Die Ergebnisse der zufälligen Versuchspläne sind über 100 Wiederholungen gemittelt.

Nr.	Zuordnung der Komponenten	Rotation	$D_{\text{KL}}$	$d_S$
1)	zufällig	keine	10.7	0.82
2)	zufällig		4.3	0.79
3)	lokale Suche		3.8	0.79
4)	Punktselektion nach lokaler Suche		3.1	0.79
5)	Kombination Punktselektion & lokale Suche		<b>2.7</b>	<b>0.74</b>

beschriebene Punktselektion angewandt, wodurch die Gleichförmigkeit weiter zunimmt. Indem ähnliche Versuchspunkte entfernt werden, solange das Kriteriums  $d_S$  unverändert bleibt, ergibt sich in 4) ein auf  $N_{\text{Sub}} = 38$  Punkte reduzierter Versuchsplan. Schließlich kann durch die Kombination von lokaler Suche und Punktselektion in 5) mit ebenfalls  $N_{\text{Sub}} = 38$  Punkten ein Versuchsplan mit den besten Ergebnissen bei beiden Kriterien erstellt werden. Dies verdeutlicht, dass das sequentielle Entfernen von Punkten durch die Punktselektion weitere Tauschvorgänge im lokalen Suchalgorithmus ermöglicht und somit zu einem besseren Optimierungsergebnis führen kann.

### Versuchsdurchführung und Modellbildung

Aufgrund der besten Bewertung wurden die realen Versuche mit dem Versuchsplan 5) durchgeführt. Da die Montageanlage für den betrachteten Trommeltyp noch nicht vollständig in Betrieb genommen war, wurden alle Komponenten manuell einander zugeordnet und der Montagevorrichtung zugeführt. Alle Trommeln wurden anschließend halbautomatisch vermessen, wobei eine Trommel aufgrund einer fehlerhaften Messung nicht in den Trainingsdatensatz aufgenommen werden konnte. Die durch die Punktselektion aussortierten Komponenten wurden zu  $N - N_{\text{Sub}} = 7$  zusätzlichen Baugruppen montiert, um einen unabhängigen Testdatensatz zu erhalten. Hierzu wurde ein gesonderter raumfüllender Versuchsplan mittels lokaler Suche erstellt.

Auf Basis der Versuchsdaten werden zwei unabhängige Gaußprozessmodelle (GPM) erstellt, die aus den vier betrachteten Eingangsgrößen jeweils eine Komponente der

radialen Abweichung  $\hat{r}_{x,T}$  bzw.  $\hat{r}_{y,T}$  vorhersagen. Zum Vergleich wird mit

$$[\hat{r}_{x,T} \ \hat{r}_{y,T}]^T = [r_{x,1} + r_{x,2} \ r_{y,1} + r_{y,2}]^T \quad (5.3)$$

ein einfaches theoretisches Modell (TM) betrachtet, das die radialen Abweichungen der einzelnen Komponenten addiert. Eine solche Addition kann angenommen werden, wenn der Einfluss des Montageprozesses vernachlässigbar ist, sodass die Abweichungen der Komponenten unverändert in die Baugruppen einfließen.

Die gemessenen und die vorhergesagten Ausgangsgrößenvektoren werden in Bild 5.15 anhand ihrer Norm verglichen. Dies erfolgt in (a) auf Basis der Trainingsdaten und in (b) auf Basis der Testdaten. In beiden Bildern sind deutliche Abweichungen des TM erkennbar, die durch eine Vernachlässigung der Verformung von Komponenten bei der Montage erklärt werden können. Ein auf Basis der Versuchsdaten trainiertes Modell wie das GPM ist in der Lage, diese Einflüsse abzubilden, was anhand der höheren Konzentration von Punkten entlang der gestrichelten Linie erkennbar ist.

Zusätzlich wird ein Polynommodell betrachtet, das mittels *Stepwise Regression* (SR) erstellt wurde. Durch die SR werden nur statistisch signifikante Regressoren eines Polynommodells in das Regressionsmodell aufgenommen, was zu einer Verringerung der Parameteranzahl führt und die Modellqualität steigert [80]. Mittels SR werden in dieser Anwendung quadratische Polynommodelle mit vier Eingangsgrößen und

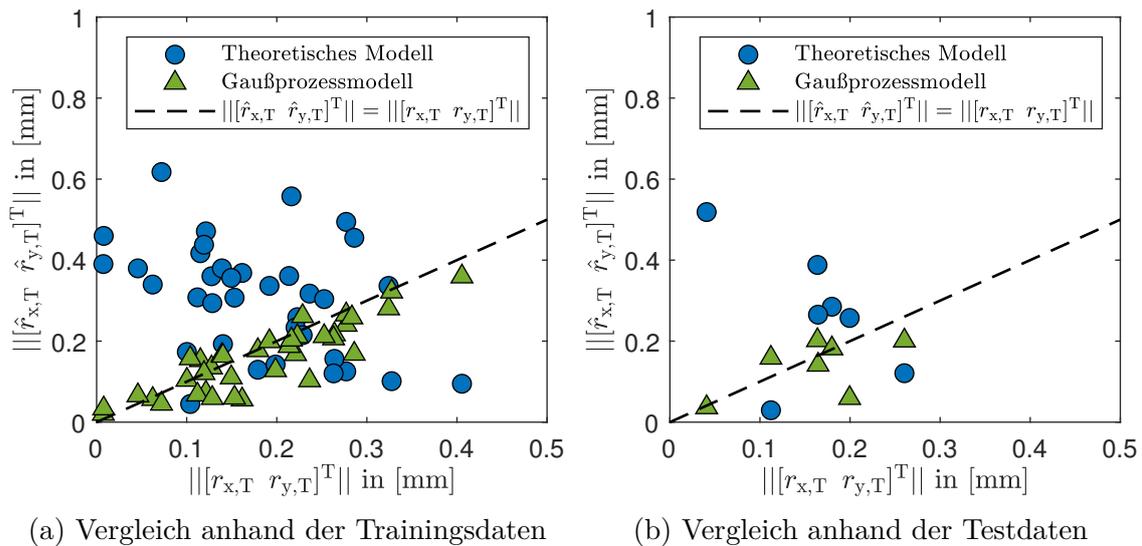


Bild 5.15: Vergleich der gemessenen und vorhergesagten Abweichungen der montierten Trommeln anhand der Vektornormen [115]. Entlang der gestrichelten Linie sind vorhergesagte und gemessene Norm identisch.

ursprünglich 15 Parametern auf sechs ( $r_{x,T}$ ) bzw. fünf ( $r_{y,T}$ ) signifikante Parameter reduziert.

Tabelle 5.5 fasst die mit den verschiedenen Modellen erzielten Ergebnisse zusammen. Beide auf den Versuchsdaten basierenden Modelle SR und GPM weisen gegenüber dem TM einen geringeren Fehler auf. In Bezug auf die verschiedenen Richtungen der Abweichungen zeigt sich, dass die x-Richtung ( $r_{x,T}$ ) von beiden datenbasierten Modellen besser vorhergesagt werden kann als die y-Richtung ( $r_{y,T}$ ). Würden die Abweichungen der Baugruppen ausschließlich aus den betrachteten Komponenten resultieren, wären aufgrund der radialsymmetrischen Komponenten ähnliche Fehler für beide Richtungen zu erwarten. Es kann daher geschlossen werden, dass diese Asymmetrie durch Einflüsse der Montageanlage hervorgerufen wird.

Für die Qualität einer Baugruppe ist die Gesamtabweichung ausschlaggebend, die durch die Norm beschrieben wird. Hierbei erzielt das SR-Modell den geringsten Fehler. Aus diesem Grund und durch die im Vergleich zum GPM einfachere Implementierung in der Anlagensteuerung der Montageanlage wird dieses Modell für den praktischen Einsatz bevorzugt. Die Validierung des Modells im realen Betrieb steht aufgrund der noch nicht abgeschlossenen Inbetriebnahme der Montageanlage während der Erstellung dieser Arbeit noch aus.

## 5.4 Zusammenfassung

In diesem Kapitel wurden drei verschiedene Anwendungen für die entwickelten Versuchsplanungsmethoden vorgestellt, die zur Demonstration der praktischen Anwendbarkeit dienen. Die auf Basis realer Datensätze erstellten Versuchspläne und Modelle wurden bewertet sowie deren Nutzen in den Anwendungen aufgezeigt.

Tabelle 5.5: Bewertung der Modellgüte für die radiale Abweichung in x-Richtung  $r_{x,T}$ , y-Richtung  $r_{y,T}$  und die Norm  $\|[r_{x,T} \ r_{y,T}]^T\|$ . Das theoretische (TM), das Stepwise Regression (SR) und das Gaußprozessmodell (GPM) werden mittels RMSE anhand des Testdatensatzes bewertet.

Modell	$r_{x,T}$ in [mm]	$r_{y,T}$ in [mm]	$\ [r_{x,T} \ r_{y,T}]^T\ $ in [mm]
TM	0.3187	0.2210	0.2170
SR	0.1126	<b>0.1470</b>	<b>0.0557</b>
GPM	<b>0.1024</b>	0.1845	0.0625

Die inkrementelle Modellbildung wurde zuerst auf ein Labormodell einer Heizstrecke angewandt. An diesem Labormodell konnten im Gegensatz zu realen industriellen Prozessen mehrere Versuchspläne und umfangreiche Testdatensätze vermessen werden. Mit der inkrementellen Modellbildung wurde eine vergleichbare bzw. in der Nähe des Arbeitspunkts sogar höhere Modellgüte erzielt als mit einem klassischen Vorgehen zur experimentellen Modellbildung. Insbesondere konnte anhand dieser Anwendung der praktische Nutzen der Transformation von Versuchsplänen in Verbindung mit lokalen Modellnetzen demonstriert werden, die auf der Partitionierung des Eingangsraums in vorherigen Schritten beruht. Eine solche Nutzung von Informationen aus vorherigen Schritten und der verminderte Stellaufwand bei der Versuchsdurchführung wurden als wesentliche Vorteile der Methodik gegenüber klassischen Verfahren deutlich.

Als zweite Anwendung diente ein Bitumenofen, der mittels der inkrementellen Modellbildung vermessen und modelliert wurde. Hierbei wurden Randbedingungen berücksichtigt, die den zulässigen Eingangsraum einschränken. Auf Basis der Versuchsdaten wurden verschiedene inkrementelle Modelle erstellt und anhand eines zusätzlich aufgenommenen Testdatensatzes verglichen. Die bezogen auf die Modellqualität besten Ergebnisse erzielte das Gaußprozessmodell GPM-DP. Schließlich wurde eine Optimierungsstrategie entwickelt und in der Anlagensteuerung des Bitumenofens implementiert, die ein inkrementelles Modell zur Bestimmung optimaler Betriebsparameter nutzt. Im regulären Produktionsbetrieb resultieren hieraus geringere Abweichungen der Bitumenmattentemperatur vom vorgegebenen Sollwert als mit den zuvor manuell festgelegten Arbeitspunkten. Auf diese Weise wird die mit dem Prozess erzielte Qualität gesteigert und zugleich der Energiebedarf des Ofens verringert.

Als Anwendung für die entwickelte Methodik zur Versuchsplanung für zusammengesetzte Einheiten wurde ein Montageprozess von Waschmaschinentrommeln betrachtet. Anhand von zwei unterschiedlichen Trommeltypen mit separaten Montageanlagen konnten zwei verschiedene Versuchsplanungsansätze für die Berücksichtigung der Rotation von Komponenten angewandt werden. Beim lokalen Ansatz wird ein Modell verwendet, das Abweichungen an einer lokalen Position aus Merkmalen an dieser Positionen vorhersagt, aber dabei Einflüsse anderer Positionen vernachlässigt. Mit diesem Ansatz konnten Höhenabweichungen eines Trommeltyps vorhergesagt werden. Der globale Ansatz verwendet hingegen Merkmale, die eine mittlere Abweichung der verwendeten Komponenten beschreiben. Hiermit konnten Rundlaufabweichungen eines anderen Trommeltyps vorhergesagt werden. Die erstellten Modelle bieten die Möglichkeit, im regulären Prozessbetrieb zu einer Verbesserung der Qua-

lität der Waschmaschinentrommeln beizutragen, indem gegebene Komponenten einander zugeordnet oder rotatorisch so ausgerichtet werden, dass die vorhergesagten Abweichungen minimal sind.

## 6 Zusammenfassung und Ausblick

In dieser Arbeit wurden zwei Methoden zur experimentellen Modellbildung entwickelt, die mit einem geringen Aufwand auf bestehende mechatronische Prozesse angewendet werden können. Die resultierenden Modelle eignen sich insbesondere, um die betrachteten Prozesse zu optimieren, sodass die Produktqualität gesteigert und der Ressourcen- bzw. Energiebedarf verringert werden kann. Bei der Modellbildung stehen Aufwände bei der Versuchsplanung und -durchführung im Fokus, die durch eine effiziente Nutzung von Prozesswissen und die Ausnutzung von typischen Prozesseigenschaften möglichst klein ausfallen.

### **Inkrementelle Modellbildung**

Die experimentelle Modellbildung wird mit steigender Dimensionalität des Eingangsraums komplexer, sodass Versuche aufwendiger zu planen, durchzuführen und zu überwachen sind. Mit der inkrementellen Modellbildung wird in dieser Arbeit ein schrittweises Verfahren vorgestellt, das mit einer Dimensionalität von eins beginnt und mit jedem Schritt eine weitere Eingangsgröße einbezieht. Somit ist ein intuitives Vorgehen mit ansteigender Komplexität gegeben. Durch eine im Vorfeld erstellte Rangfolge werden die wichtigsten Eingangsgrößen zuerst betrachtet, was eine Abbildung des wesentlichen Prozessverhaltens in wenigen Schritten erlaubt. Der inkrementelle Versuchsplan, der auf Basis von optimierten Latin Hypercubes erstellt wird und auf einen vorgegebenen Arbeitspunkt ausgerichtet ist, sowie die inkrementelle Modellstruktur, bestehend aus additiv überlagerten Teilmodellen, werden in jedem Schritt um die weitere Eingangsgröße ergänzt.

Als Teilmodellstrukturen werden Polynommodelle, lokale Modellnetze (LMN) und Gaußprozessmodelle (GPM) betrachtet, die in dieser Arbeit so erweitert werden, dass die bereits erzielte Modellgüte in weiteren Modellierungsschritten nur gesteigert werden kann. Die inkrementellen Modelle mit allen betrachteten Teilmodellstrukturen sind in der Lage, eine ähnliche oder besser Modellgüte in der Nähe des vorgegebenen

Arbeitspunkts zu erzielen, wie ein nicht inkrementelles Vergleichsmodell derselben Modellklasse. Anhand der Polynommodelle werden die Vorteile raumfüllender Versuchspläne demonstriert, wenn die exakten funktionalen Zusammenhänge des Prozesses unbekannt sind. In solchen Fällen können höhere Modellgüten erzielt werden als mit optimalen Versuchsplänen, die auf Fehlannahmen über die funktionalen Zusammenhänge beruhen. Durch das schrittweise Vorgehen kann das bereits modellierte Prozessverhalten in die Versuchsplanung in folgenden Schritten einfließen. Hierzu wird eine Transformation des inkrementellen Versuchsplans auf Grundlage der Partitionierung von LMN in den vorherigen Schritten beschrieben, um diesen an bereits detektierte Nichtlinearitäten des Prozesses anzupassen. Darüber hinaus ermöglicht die inkrementelle Modellstruktur zusätzliche Interpretationen wie die Detektion von Haupteffekten und Wechselwirkungen. Hierzu werden die Standardabweichungen der in einem GPM verwendeten Kernel-Funktionen ausgewertet.

Die praktische Anwendbarkeit der inkrementellen Modellbildung wird anhand einer Heizstrecke und eines Bitumenofens demonstriert. Hierbei werden an der Heizstrecke verschiedene inkrementelle Versuchspläne verglichen. Für den Bitumenofen wird die Realisierung einer modellbasierten Optimierung auf Basis eines inkrementellen Modells beschrieben.

### **Versuchsplanung für zusammengesetzte Einheiten**

Die Eigenschaften von zusammengesetzten Einheiten werden durch die in den Einheiten verwendeten Komponenten beeinflusst. Jede Komponente kann durch eine oder mehrere Eingangsgrößen beschrieben werden, die beispielsweise deren Qualität charakterisieren und für die sie feste Werte vorgibt. Ziel der Versuchsplanung für zusammengesetzte Einheiten ist es, eine bestmögliche Datenverteilung aus gegebenen Sätzen an Komponenten zu erzeugen, wobei raumfüllende Versuchspläne angestrebt werden. Hierzu werden mit einem lokalen Suchalgorithmus und einem genetischen Algorithmus zwei verschiedene Optimierungsalgorithmen betrachtet, die optimale Kombinationen der Komponenten erzeugen. Mit dem lokalen Suchalgorithmus werden im Rahmen dieser Arbeit bessere raumfüllende Eigenschaften erzielt.

Durch ein geändertes Optimierungskriterium können dieselben Algorithmen zur Erzeugung möglichst ähnlicher Versuchspunkte verwendet werden, die als näherungsweise Wiederholpunkte dienen. Indem einige näherungsweise Wiederholpunkte in

einem ansonsten raumfüllenden Versuchsplan platziert werden, können bei vergleichbarer Modellgüte Rauscheinflüsse besser geschätzt werden. Dieses Verhalten wird in Verbindung mit GPM demonstriert und führt zu einer gesteigerten Interpretierbarkeit der Modelle. Trotz der Optimierung können aus realen Datenverteilungen zahlreiche Punkte resultieren, die nah beieinanderliegen. Die Vermessung dieser Punkte verursacht Aufwand in den Versuchen, liefert allerdings wenig Informationen über den Prozess. Durch die Anwendung eines Verfahrens zur Punktselektion hat sich gezeigt, dass bei solchen Datenverteilungen viele Punkte entfernt werden können, ohne die Modellgüte signifikant herabzusetzen.

Die Einbeziehung der Rotation von Komponenten stellt eine zusätzliche Erweiterung dar, die bei Montageprozessen von radialsymmetrischen Baugruppen genutzt werden kann. Für die Versuchsplanung und Modellbildung werden dabei mit einem globalen und einem lokalen Ansatz zwei unterschiedliche Ansätze für Auswirkungen der Komponenten auf die Baugruppe betrachtet. Indem Komponenten in veränderten Orientierungen verbaut werden, ergeben sich größere Freiheiten in der Versuchsplanung, die zu gleichmäßigeren Datenverteilungen führen. Zur Verifizierung der entwickelten Versuchsplanungsmethodik unter Einbeziehung beider Ansätze zur Rotation von Komponenten werden zwei verschiedene Montageprozesse für Waschmaschinentrommeln als reale Anwendungen betrachtet. Die resultierenden Modelle dienen zur Vorhersage von Qualitätsgrößen und können somit zur Verbesserung der Montageprozesse eingesetzt werden.

## **Ausblick**

Folgende Themenfelder können im Rahmen zukünftiger Forschungsvorhaben adressiert werden, um ein breiteres Anwendungsfeld der entwickelten Methoden zu ermöglichen oder deren praktische Anwendbarkeit zu steigern:

*Einsatz sequentieller Versuchspläne:* Die Festlegung einer Anzahl an Versuchspunkten im Vorfeld der Messungen stellt eine große Herausforderung in der Praxis dar. Sequentielle Versuchspläne erfordern keine solche Festlegung und ermöglichen mehrere kleine Experimente nacheinander anstatt eines einzelnen umfangreichen Experiments. Zugleich können Informationen aus vorherigen Messungen in die weitere Versuchsplanung einbezogen werden [60]. Indem sequentielle Versuchspläne sowohl in jedem Schritt der ILHAD-Methodik als auch in der Versuchsplanung für zusam-

mengesetzte Einheiten verwendet werden, können Versuchsressourcen effizienter eingesetzt werden.

*Detektion des zulässigen Eingangsraums:* Mit steigender Dimensionalität des Eingangsraums ist eine manuelle Beschreibung der multivariaten Grenzen von zulässigen Bereichen zunehmend schwieriger. Automatische Methoden zur Erkundung des zulässigen Eingangsraums [93] können dazu beitragen, diese Aufgabe zu vereinfachen. Eine solche Kombination wäre sowohl für ILHAD als auch für die Versuchsplanung für zusammengesetzte Einheiten eine sinnvolle Erweiterung, die zu einer sicheren Versuchsdurchführung beitragen kann.

*Dynamische Modellbildung:* Die ILHAD-Methode kann auch auf dynamische Prozesse angewandt werden, indem die einzelnen dynamischen Eingangsgrößen schrittweise einbezogen und angeregt werden. Neben geeigneten Anregungssignalen werden hierfür dynamische Modelle benötigt, die sich für den Einsatz in der additiven Modellstruktur von ILHAD eignen.

# A Anhang

## A.1 Grundlagen des EDLS-Algorithmus

Bei dem Extended Deterministic Local Search-Algorithmus (EDLS) [26] handelt es sich um einen lokalen Suchalgorithmus, der zur Optimierung von Latin Hypercubes (LHC) entwickelt wurde. Das verwendete Optimierungskriterium basiert auf dem Maximin-Kriterium (siehe Gleichung 2.28) und wird durch wiederholte Koordinatentauschvorgänge zwischen den Versuchspunkten verbessert. Auf diese Weise ermöglicht der Algorithmus die Erstellung von raumfüllenden Versuchsplänen mit einem moderaten Rechenaufwand.

Der grundlegende Ablauf des EDLS-Algorithmus ist in Bild A.1 dargestellt. In (a) wird der Versuchsplan  $\mathcal{D}(0) = \mathcal{D}$  mit einem vorgegebenen Versuchsplan  $\mathcal{D}$  mit  $N$  Versuchspunkten initialisiert. Zusätzlich wird eine Dimension für den Koordinatentausch  $j$  und eine Menge an ausgeschlossenen Punkten mit  $\mathcal{E} = \emptyset$  initialisiert. Zu Beginn jeder Iteration  $i \in \mathbb{N}$  wird in (b) der kritische Punkt  $\underline{u}_{\text{krit}}$  und dessen nächster Nachbar  $\underline{u}_{\text{NN}}$  identifiziert. Der kritische Punkt bezeichnet einen der Versuchspunkte, die zu der geringsten Nächsten-Nachbarn-Distanz

$$\tilde{d}_{\min} = \min_{\substack{\underline{u}_{\text{krit}} \in \mathcal{D}(i) \setminus \mathcal{E}, \\ \underline{u}_{\text{NN}} \in \mathcal{D}(i), \underline{u}_{\text{krit}} \neq \underline{u}_{\text{NN}}}} \left( d(\underline{u}_{\text{krit}}, \underline{u}_{\text{NN}}) \right), \quad (\text{A.1})$$

führen, wobei derjenige Punkt mit dem niedrigeren Index als  $\underline{u}_{\text{krit}}$  gewählt wird. In (c) werden anschließend die möglichen Tauschpartner  $\mathcal{M} = \mathcal{D}(i) \setminus (\{\underline{u}_{\text{krit}}, \underline{u}_{\text{NN}}\} \cup \mathcal{E})$  für einen Koordinatentausch ermittelt, aus denen in (d) jeweils ein neuer Tauschpartner  $\underline{u}^*$  ausgewählt wird. Der Koordinatentausch erfolgt in (e) zwischen dem kritischen Punkt  $\underline{u}_{\text{krit}}$  und dem Tauschpartner  $\underline{u}^*$  in der Dimension  $u_j$ , sodass sich ein neuer Versuchsplan  $\mathcal{D}^*$  mit der Nächsten-Nachbarn-Distanz

$$\tilde{d}_{\min}^* = \min_{\substack{\underline{u}_1 \in \mathcal{D}^* \setminus \mathcal{E}, \\ \underline{u}_2 \in \mathcal{D}^*, \underline{u}_1 \neq \underline{u}_2}} \left( d(\underline{u}_1, \underline{u}_2) \right) \quad (\text{A.2})$$

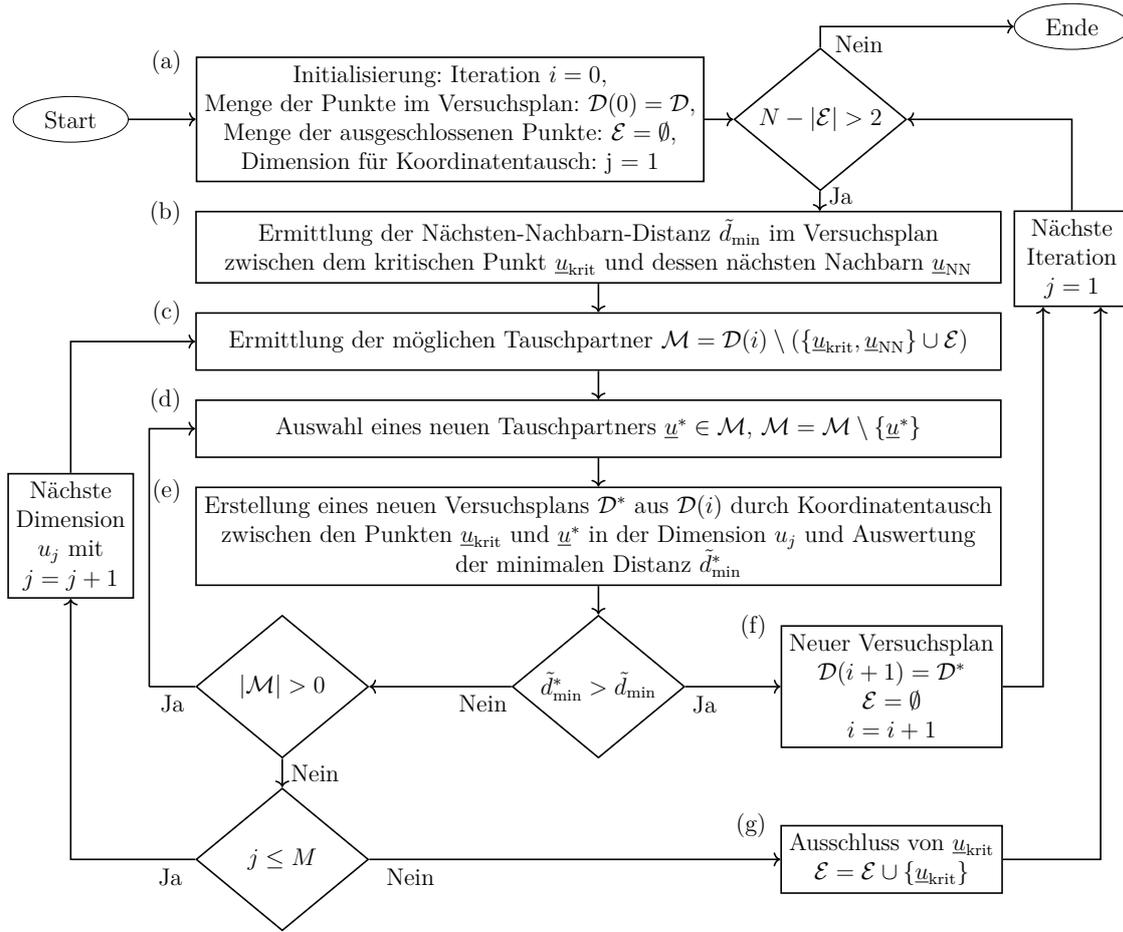


Bild A.1: Grundlegender Ablauf des EDLS-Algorithmus.

ergibt. Eine Vergrößerung der Nächsten-Nachbarn-Distanz durch den Tausch fungiert als Tauschbedingung

$$\tilde{d}_{\min}^* > \tilde{d}_{\min} \quad (\text{A.3})$$

des EDLS-Algorithmus. Ist diese Bedingungen erfüllt, beginnt eine neue Iteration mit dem Versuchsplan  $\mathcal{D}(i+1) = \mathcal{D}^*$ . Ist die Tauschbedingung nicht erfüllt, werden systematisch weitere Koordinatentauschvorgänge überprüft.

Wurden mit  $|\mathcal{M}| = 0$  alle Tauschvorgänge mit allen möglichen Tauschpartnern in allen Dimensionen  $u_j$  mit  $j = 1, 2, \dots, M$  überprüft, ist keine Optimierung mit dem in der jeweiligen Iteration ermittelten kritischen Punkt  $\underline{u}_{\text{krit}}$  möglich. Dieser Punkt wird daher in die Menge  $\mathcal{E}$  aufgenommen und ist so von den nachfolgenden Koordinatentauschvorgängen ausgeschlossen. Dieser Ausschluss von Punkten ermöglicht es dem EDLS-Algorithmus nicht nur die kleinste Nächste-Nachbarn-Distanz gemäß Gleichung (2.21), sondern auch weitere Distanzen wie die zweit- oder drittkleinste Punktdistanz im Versuchsplan zu maximieren. Da ein erfolgreich durchgeführter

Koordinatentausch neue Tauschvorgänge für bereits ausgeschlossene Punkte ermöglichen kann, wird die Menge der ausgeschlossenen Punkte in (f) auf  $\mathcal{E} = \emptyset$  zurückgesetzt. Der Algorithmus wird beendet, wenn  $N - |\mathcal{E}| = 2$  gilt, sodass in (c) mit  $\mathcal{D}(i) \setminus (\{u_{\text{krit}}, u_{\text{NN}}\} \cup \mathcal{E}) = \emptyset$  keine Tauschpartner mehr zur Verfügung stehen.

## A.2 Berücksichtigung von Randbedingungen im EDLS-Algorithmus

Viele reale Prozesse weisen Betriebsbereiche auf, die zu einem unzulässigen Prozessverhalten führen und daher bei der Durchführung von Versuchen vermieden werden müssen. Ein Beispiel ist der in Abschnitt 5.2 betrachtete Bitumenofen, bei dem eine unzulässige Erwärmung der Bitumenmatten vermieden werden muss. Um solche Prozesse zu vermessen, müssen diese unzulässigen Bereiche des Eingangsraums in Form von Randbedingungen bei der Versuchsplanung berücksichtigt werden.

Im Folgenden wird der zulässige Eingangsraum durch eine beliebige mathematische Funktion beschrieben, die einen Wahrheitswert liefert. Als Beispiel wird in einem zweidimensionalen Eingangsraum die Randbedingung

$$u_2 \geq 0.4 \cdot u_1 - 0.1 \wedge u_2 \leq 0.6 \cdot u_1 + 0.7 \quad (\text{A.4})$$

verwendet, die sich aus zwei linearen Ungleichungen zusammensetzt. Zur Berücksichtigung dieser Randbedingung im EDLS-Algorithmus [26] wird die Bedingung für einen akzeptierten Koordinatentausch erweitert. Nur wenn beide Punkte, die sich durch einen Koordinatentausch ergeben, zur Verbesserung des Versuchsplans führen (Gleichung (A.3)) und zusätzliche die Randbedingung (Gleichung (A.4)) erfüllen, wird der Tauschvorgang durchgeführt.

Bild A.2 stellt in (a) einen Versuchsplan mit drei Punkten und der resultierenden minimalen Distanz  $d_{\text{min}}$  dar, die optimiert werden soll. Hierzu werden gemäß dem EDLS-Algorithmus zwei verschiedene Koordinatentauschoperationen in der Dimension  $u_1$  dargestellt. Beide Tauschvorgänge führen zu einer Vergrößerung der minimalen Distanz  $d_{\text{min}}$  und damit zu einem verbesserten Versuchsplan im Sinne des Optimierungskriteriums. Unter Berücksichtigung der Randbedingung ist allerdings nur der Tauschvorgang (c) zulässig, da bei (b) ein Punkt außerhalb des zulässigen Eingangsraums liegen würde.

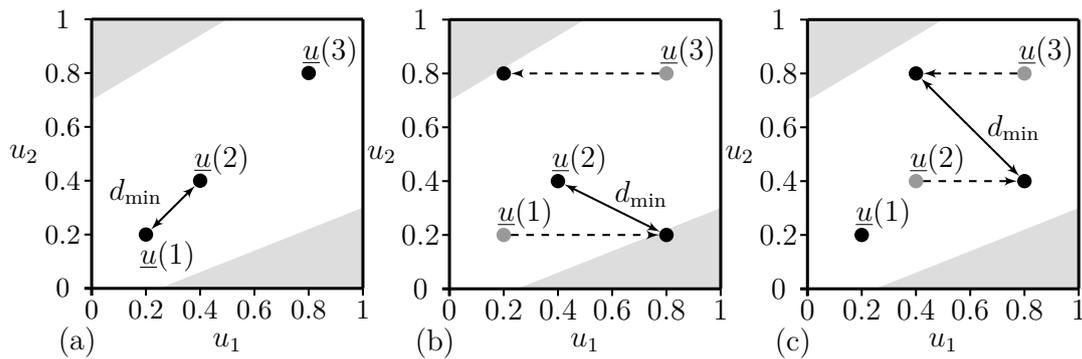


Bild A.2: Optimierung eines Versuchsplans durch den EDLS-Algorithmus unter Berücksichtigung von Randbedingungen. Durch Koordinatentausch werden die Punkte von ihren vorherigen Positionen (grau) verschoben. Der Tauschvorgang in (b) ist durch die Randbedingungen unzulässig.

Bild A.3 stellt einen auf diese Weise optimierten Versuchsplan für zwei Eingangsgrößen dar. Im ersten Schritt wird ein initialer Versuchsplan (a) erzeugt, dessen Punkte ausschließlich im zulässigen Eingangsraum liegen. Die Optimierung dieses Versuchsplans erfolgt mittels des EDLS-Algorithmus unter Berücksichtigung der Randbedingungen und führt zu dem raumfüllenden Versuchsplan in (b).

Bei einem großen Anteil der unzulässigen Bereiche am gesamten Eingangsraum kann das beschriebene Verfahren zur Optimierung eines Versuchsplans zu einer ungenügenden Raumabdeckung führen. So werden die Möglichkeiten zum Koordinatentausch durch die Randbedingungen eingeschränkt, womit die Zahl lokaler Optima steigt. Beispielsweise könnte ein Versuchsplan ohne Beachtung der Randbedingung durch die Durchführung von zwei aufeinanderfolgenden Tauschoperationen verbessert werden und anschließend trotzdem die Randbedingung wieder erfüllen. Da im beschriebenen Algorithmus jedoch jede Verletzung der Randbedingung unterbunden wird, kann ein solcher Versuchsplan nicht durch die beschriebene Strategie erzeugt werden. Bei großen unzulässigen Bereichen im Eingangsraum könnten bessere Ergebnisse mit Verfahren wie in [85] erzielt werden, bei denen Randbedingungen zu Beginn verletzt und die Einhaltung durch wiederholte Tauschoperationen herbeigeführt wird.

## A.3 Grundlagen genetischer Optimierung

Genetische Algorithmen zählen zu den evolutionären Algorithmen und bilden damit evolutionäre Prozesse nach. Hierbei werden  $A$  Individuen  $S_l$  mit  $l = 1, 2, \dots, A$

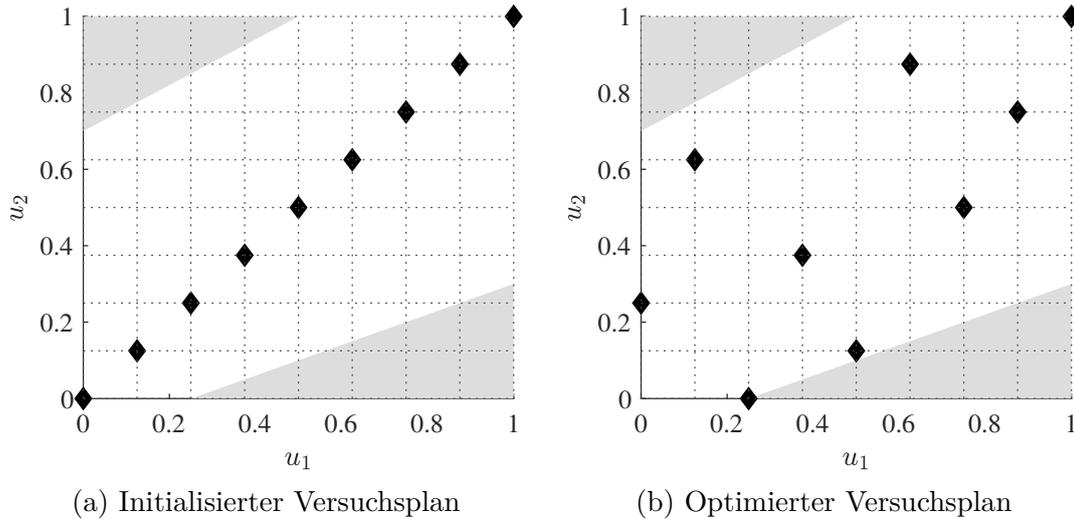


Bild A.3: Versuchsplan für zwei Eingangsgrößen mit Randbedingungen vor und nach der Optimierung mittels EDLS-Algorithmus. Die grau hinterlegten Bereiche im Eingangsraum wurden als unzulässig definiert.

betrachtet, die jeweils durch eine Gensequenz codierte Eigenschaften besitzen. Diese Eigenschaften repräsentieren die zu optimierenden Parameterwerte und werden mittels eines Optimierungskriteriums bewertet. Die Individuen werden zu einer Generation  $\mathcal{G} = \{S_1, S_2, \dots, S_A\}$  zusammengefasst. In jedem Iterationsschritt (*Epoche*)  $j$ , mit  $j = 1, 2, \dots, W_G$  des genetischen Algorithmus wird aus den Individuen eine Generation  $\mathcal{G}(j)$  eine Nachfolgegeneration  $\mathcal{G}(j + 1)$  mit veränderten Gensequenzen erzeugt. Hierbei wird durch die Verwendung verschiedener genetischer Operationen eine Verbesserung des Optimierungskriteriums angestrebt, aber auch durch Einbringung zufälliger Variationen ein Verlassen lokaler Optima ermöglicht [118].

Bild A.4 stellt die grundlegenden Elemente des in dieser Arbeit verwendeten genetischen Algorithmus und dessen Ablauf dar. Zu Beginn wird in (a) eine Startgeneration  $\mathcal{G}(1)$  erstellt, deren Individuen zufällig initialisiert werden. Die Nachfolgegeneration wird in (b) als leere Menge initialisiert. Für jedes Individuum in der aktuellen Generation  $\mathcal{G}(j)$  wird anschließend in (c) die Fitnessfunktion  $F(\cdot)$  ausgewertet. Mittels dieser Fitnessfunktion werden die einzelnen Individuen bewertet. Individuen mit einer höheren Bewertung werden bevorzugt zur Erstellung der Nachfolgegeneration verwendet. Trotzdem soll die Vielfalt in der Nachfolgegeneration nicht zu sehr eingeschränkt werden. Die Fitnessfunktion basiert auf der zu optimierenden Zielfunktion, die beispielsweise so skaliert wird, dass Unterschiede zwischen den Individuen verstärkt werden [118]. In (d) werden basierend auf den Fitnesswerten jeweils zwei Individuen  $S_A$  und  $S_B$  aus der aktuellen Generation  $\mathcal{G}(j)$  selektiert. Die Selektions-

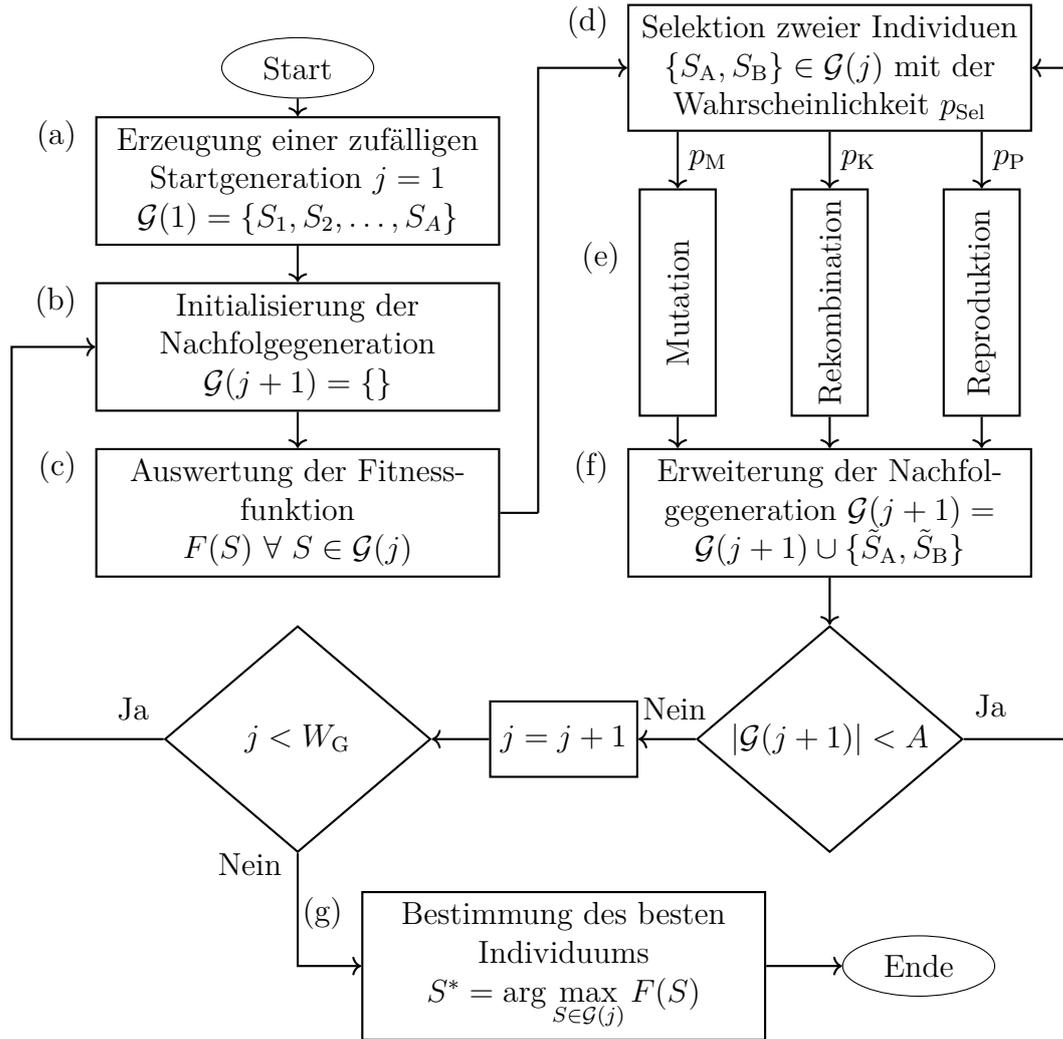


Bild A.4: Grundlegender Ablauf der genetischen Optimierung.

wahrscheinlichkeit eines Individuums

$$p_{\text{Sel}}(S) = \frac{F(S)}{\sum_{i=1}^A F(S_i)} \quad (\text{A.5})$$

wird dabei üblicherweise proportional zu dessen Fitnesswert  $F(S)$  gewählt. Dieser Selektionsmechanismus wird *Roulette Wheel Selection* genannt und kann mit dem Drehen eines Rouletterads veranschaulicht werden, dessen Flächen proportional zu den Fitnesswerten der einzelnen Individuen sind [80]. Beide Individuen werden in (e) gemeinsam einer zufällig ausgewählten genetischen Operation unterzogen, wobei die Wahrscheinlichkeiten  $p_M$  für eine Mutation,  $p_K$  für eine Rekombination und  $p_P$  für eine Reproduktion gegeben sind. Diese drei genetischen Operationen haben folgende Auswirkungen auf die Individuen:

- Bei einer Mutation werden zufällige Veränderungen in den Gensequenzen der jeweiligen Individuen durchgeführt, die das Verlassen lokaler Optima ermöglichen. Hierbei werden das Individuum  $\tilde{S}_A$  durch Veränderung von  $S_A$  und  $\tilde{S}_B$  durch Veränderung von  $S_B$  erzeugt. Bei der Mutation werden typischerweise kleine Änderungen an den Individuen vorgenommen, wie die Veränderung eines einzelnen zufällig ausgewählten Bits in der Gensequenz [59].
- Ein Rekombinationsvorgang erzeugt zwei neue Individuen  $\tilde{S}_A$  und  $\tilde{S}_B$ , die eine Kombination der Eigenschaften von  $S_A$  und  $S_B$  darstellen. Jedes der neuen Individuen erhält dabei Teile der Gensequenzen der vorherigen Individuen. Durch die gleichzeitige Erzeugung zweier nachfolgender Individuen bestehen alle Teile der Gensequenzen weiter fort, sodass hierdurch in der gesamten Generation kein Informationsverlust eintritt [118].
- Die Reproduktion fügt Individuen der aktuellen Generation unverändert in die Nachfolgeneraion ein, sodass  $\tilde{S}_A = S_A$  und  $\tilde{S}_B = S_B$  gilt. Durch die Abhängigkeit der Selektionswahrscheinlichkeit vom Fitnesswert (Gleichung A.5) werden die besten Individuen bevorzugt reproduziert [37]. Gute Lösungen können mittels dieser Operation fortbestehen.

Die aus einer der genetischen Operationen resultierenden Individuen werden anschließend in (f) zur Nachfolgeneraion hinzugefügt. Die Schritte (d) bis (f) werden solange iterativ durchlaufen, bis dieselbe Anzahl  $A$  an Individuen in der Nachfolgeneraion enthalten ist wie in der Startgeneraion.

Ist eine Nachfolgeneraion erstellt, ist eine Epoche des genetischen Algorithmus abgeschlossen. Typischerweise wird eine festgelegte Anzahl an Epochen  $W_G$  durchlaufen, wobei die Schritte (b) bis (f) wiederholt ausgeführt werden. Schließlich wird in (g) aus der aktuellen Generaion das Individuum mit dem höchsten Fitnesswert als Lösung des Optimierungsproblems  $S^*$  ausgewählt.

Für genetische Algorithmen existieren zahlreiche Variationsmöglichkeiten, die beispielsweise die Berechnung der Fitnesswerte, die Realisierung der genetischen Operationen oder die Abbruchbedingung betreffen und eine Anpassung auf verschiedene Optimierungsprobleme ermöglichen. Weitere Informationen hierzu sind in den Büchern [37, 59, 118] zu finden.



---

## Literaturverzeichnis

- [1] Akcil, A., Sun, Z., Panda, S.: COVID-19 disruptions to tech-metals supply are a wake-up call. *Nature* **587**(7834), 365–367 (2020). DOI 10.1038/d41586-020-03190-8
- [2] Anderson-Cook, C.M., Robinson, T.J.: A Designed Screening Study with Pre-specified Combinations of Factor Settings. *Quality Engineering* **21**(4), 392–404 (2009). DOI 10.1080/08982110903179069
- [3] Ankenman, B., Nelson, B.L., Staum, J.: Stochastic kriging for simulation metamodeling. In: 2008 Winter Simulation Conference. IEEE (2008). DOI 10.1109/wsc.2008.4736089
- [4] Anthony, D.K., Keane, A.J.: Genetic algorithms for design of experiments on assembled products. Tech. rep., University of Southampton (2004)
- [5] Arboretti, R., Ceccato, R., Pegoraro, L., Salmaso, L.: Design choice and machine learning model performances. *Quality and Reliability Engineering International* (2022). DOI 10.1002/qre.3123
- [6] Asprion, N., Böttcher, R., Pack, R., Stavrou, M.E., Höller, J., Schwientek, J., Bortz, M.: Gray-Box Modeling for the Optimization of Chemical Processes. *Chemie Ingenieur Technik* **91**(3), 305–313 (2018). DOI 10.1002/cite.201800086
- [7] Astakhov, V.P.: Screening (sieve) design of experiments in metal cutting. In: *Design of Experiments in Production Engineering*, pp. 1–37. Springer (2016)
- [8] Barbero, B.R., Aragón, A.C., Pedrosa, C.M.: Validation of a tolerance analysis simulation procedure in assemblies. *The International Journal of Advanced Manufacturing Technology* **76**(5-8), 1297–1310 (2014). DOI 10.1007/s00170-014-6229-1

- [9] Beigh, M., Hussain, S.Z., Qadri, T., Naseer, B., Raja, T., Naik, H.: Investigation of process and product parameters for physico-chemical properties of low Glycemic Index water chestnut and barley flour-based extruded snacks. *British Food Journal* **122**(1), 227–241 (2019). DOI 10.1108/bfj-01-2019-0001
- [10] Bellman, R.E.: *Adaptive Control Processes*. Princeton University Press (1961). DOI 10.1515/9781400874668
- [11] Bergmann, D., Geiselhart, R., Graichen, K.: Modelling and Control of a Heavy-Duty Diesel Engine Gas Path with Gaussian Process Regression. In: 2019 18th European Control Conference (ECC). IEEE (2019). DOI 10.23919/ecc.2019.8796016
- [12] Bergmann, D., Harder, K., Niemeyer, J., Graichen, K.: Nonlinear MPC of a Heavy-Duty Diesel Engine With Learning Gaussian Process Regression. *IEEE Transactions on Control Systems Technology* **30**(1), 113–129 (2022). DOI 10.1109/tcst.2021.3054650
- [13] Bickel, S.: *Learning under differing training and test distributions*. Ph.D. thesis, Universität Potsdam (2008)
- [14] Bisgaard, S., Weiss, P.: Experimental Design Applied to Identify Defective Parts in Assembled Products. *Quality Engineering* **15**(2), 347–350 (2002). DOI 10.1081/qen-120015867
- [15] Caffisch, R.E.: Monte Carlo and quasi-Monte Carlo methods. *Acta Numerica* **7**, 1–49 (1998). DOI 10.1017/s0962492900002804
- [16] Cao, L., Russo, D., Felton, K., Salley, D., Sharma, A., Keenan, G., Mauer, W., Gao, H., Cronin, L., Lapkin, A.A.: Optimization of Formulations Using Robotic Experiments Driven by Machine Learning DoE. *Cell Reports Physical Science* **2**(1), 100295 (2021). DOI 10.1016/j.xcrp.2020.100295
- [17] Chai, T., Draxler, R.R.: Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. *Geoscientific Model Development* **7**(3), 1247–1250 (2014). DOI 10.5194/gmd-7-1247-2014
- [18] Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pp. 785–794. Association for Computing Machinery, New York, NY, USA (2016). DOI 10.1145/2939672.2939785

- [19] Cheng, H., Chen, H.: Online parameter optimization in robotic force controlled assembly processes. In: 2014 IEEE International Conference on Robotics and Automation (ICRA). IEEE (2014). DOI 10.1109/icra.2014.6907358
- [20] Cui, S., Datcu, M.: Comparison Of Kullback-leibler Divergence Approximation Methods Between Gaussian Mixture Models For Satellite Image Retrieval. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3719–3722. IEEE (2015)
- [21] Czitrom, V.: One-Factor-at-a-Time versus Designed Experiments. *The American Statistician* pp. 126–131 (1999)
- [22] Degen, W.L.F.: *Kurven – Verstehen – Entwerfen – Berechnen – Darstellen*. KIT Scientific Publishing (2018). DOI 10.5445/KSP/1000069243
- [23] Domingos, P.: The Role of Occam’s Razor in Knowledge Discovery. *Data Mining and Knowledge Discovery* **3**(4), 409–425 (1999). DOI 10.1023/a:1009868929893
- [24] Duboue, P.: *The Art of Feature Engineering: Essentials for Machine Learning*. Cambridge University Press (2020)
- [25] Duvenaud, D., Nickisch, H., Rasmussen, C.E.: Additive Gaussian Processes. In: *Proceedings of the 24th International Conference on Neural Information Processing Systems, NIPS’11*, pp. 226–234. Curran Associates Inc. (2011)
- [26] Ebert, T., Fischer, T., Belz, J., Heinz, T.O., Kampmann, G., Nelles, O.: Extended Deterministic Local Search Algorithm for Maximin Latin Hypercube Designs. In: *IEEE Symposium Series on Computational Intelligence* (2015)
- [27] Ehrlenspiel, K., Kiewert, A., Lindemann, U., Mörtl, M.: *Kostengünstig Entwickeln und Konstruieren*. Springer Berlin Heidelberg (2014). DOI 10.1007/978-3-642-41959-1
- [28] Fahrmeir, L., Kneib, T., Lang, S.: *Regression: Modelle, Methoden und Anwendungen (Statistik und ihre Anwendungen)*. Springer (2008)
- [29] Fisher, O.J., Watson, N.J., Escrig, J.E., Witt, R., Porcu, L., Bacon, D., Rigley, M., Gomes, R.L.: Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems. *Computers & Chemical Engineering* **140**, 106881 (2020)

- 
- [30] Forsyth, D.: Applied Machine Learning. Springer International Publishing (2019). DOI 10.1007/978-3-030-18114-7
- [31] Freiesleben, J., Keim, J., Grutsch, M.: Machine learning and Design of Experiments: Alternative approaches or complementary methodologies for quality improvement? *Quality and Reliability Engineering International* **36**(6), 1837–1848 (2020). DOI 10.1002/qre.2579
- [32] Friedman, J.H.: Multivariate adaptive regression splines. *The annals of statistics* pp. 1–67 (1991)
- [33] Friedman, J.H.: Stochastic gradient boosting. *Computational statistics & data analysis* **38**(4), 367–378 (2002)
- [34] Garg, A., Tai, K., Panda, B.N.: System Identification: Survey on Modeling Methods and Models. In: *Advances in Intelligent Systems and Computing*, pp. 607–615. Springer Singapore (2017). DOI 10.1007/978-981-10-3174-8\_51
- [35] Gariani, S., El-Sayed, M.A., Shyha, I.: Optimisation of cutting fluid concentration and operating parameters based on RSM for turning Ti–6Al–4V. *The International Journal of Advanced Manufacturing Technology* **117**(1-2), 539–553 (2021). DOI 10.1007/s00170-021-07623-0
- [36] Genton, M.G.: Classes of kernels for machine learning: a statistics perspective. *Journal of machine learning research* **2**(Dec), 299–312 (2001)
- [37] Goldberg, D.E.: Genetic algorithms in search, optimization, and machine learning. Addison-Wesley Pub. Co. (1989)
- [38] Goldberg, P.W., Williams, C.K.I., Bishop, C.M.: Regression with input-dependent noise: A Gaussian process treatment. *Advances in neural information processing systems* **10**, 493–499 (1997)
- [39] Golmohamadi, H.: Demand-side management in industrial sector: A review of heavy industries. *Renewable and Sustainable Energy Reviews* **156**, 111963 (2022). DOI 10.1016/j.rser.2021.111963
- [40] Gunst, R.F., Mason, R.L.: Fractional factorial design. *Wiley Interdisciplinary Reviews: Computational Statistics* **1**(2), 234–244 (2009). DOI 10.1002/wics.27

- 
- [41] Gunzburger, M., Burkardt, J.: Uniformity measures for point sample in hypercubes. Tech. rep., Florida State University (2004). URL [https://people.sc.fsu.edu/~jburkardt/publications/gb\\_2004.pdf](https://people.sc.fsu.edu/~jburkardt/publications/gb_2004.pdf)
- [42] Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. *Journal of machine learning research* **3**(Mar), 1157–1182 (2003)
- [43] Haines, L.M.: The Application of the Annealing Algorithm to the Construction of Exact Optimal Designs for Linear – Regression Models. *Technometrics* **29**(4), 439–447 (1987). DOI 10.1080/00401706.1987.10488272
- [44] Hanitz, M.: Modellbasierte Optimierung eines Montageprozesses von Innentrommeln auf Grundlage geplanter Versuche. Masterarbeit, FH Bielefeld (2022)
- [45] Harris, P.G., O’Donnell, G.E., Whelan, T.: Modelling and identification of industrial pneumatic drive system. *The International Journal of Advanced Manufacturing Technology* **58**(9-12), 1075–1086 (2012)
- [46] Hartmann, B.: Lokale Modellnetze zur Identifikation und Versuchsplanung nichtlinearer Systeme. Ph.D. thesis, Universität Siegen (2013)
- [47] Hartmann, B., Ebert, T., Nelles, O.: Model-based design of experiments based on local model networks for nonlinear processes with low noise levels. In: *Proc. American Control Conf*, pp. 5306–5311 (2011). DOI 10.1109/ACC.2011.5990833
- [48] Hartmann, B., Nelles, O.: Identifikation mit achsenschrägen, lokal polynomia-len Modellnetzen. *at-Automatisierungstechnik* **62**(6), 394–407 (2014)
- [49] Hartmann, B., Nelles, O., Belič, A., Zupančič–Božič, D.: Local Model Networks for the Optimization of a Tablet Production Process. In: *Artificial Intelligence and Computational Intelligence*, pp. 241–250. Springer Berlin Heidelberg (2009). DOI 10.1007/978-3-642-05253-8\_27
- [50] Hickernell, F.J.: A Generalized Discrepancy and Quadrature Error Bound. *Mathematics of Computation* **67**(221), 299–322 (1998)
- [51] Hifi, M., M’Hallah, R.: A Literature Review on Circle and Sphere Packing Problems: Models and Methodologies. *Advances in Operations Research* **2009**, 1–22 (2009). DOI 10.1155/2009/150624

- [52] Horn, C., Krüger, J.: Feasibility of connecting machinery and robots to industrial control services in the cloud. In: 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFFA), pp. 1–4. IEEE (2016)
- [53] Huang, W.T., Shih, P.H., Lin, Y.C., Tsai, J.T., Chou, J.H.: Research on the Prediction Model of Milling Force and Temperature for High-Speed Milling of NAK80 Die Steel by Ultrasonic Atomization Micro-Lubrication Combined With Uniform Method and Back Propagation Neural Network. In: 2022 International Conference on Fuzzy Theory and Its Applications (iFUZZY). IEEE (2022). DOI 10.1109/ifuzzy55320.2022.9985233
- [54] Huang, Z., He, W., Zhao, L., Liu, H., Zhou, X.: Processing technology optimization for tofu curd by fermented yellow whey using response surface methodology. *Food Science & Nutrition* **9**(7), 3701–3711 (2021). DOI 10.1002/fsn3.2331
- [55] Jakob, W., Blume, C.: Pareto optimization or cascaded weighted sum: A comparison of concepts. *Algorithms* **7**(1), 166–185 (2014)
- [56] Jourdan, A., Franco, J.: Optimal Latin hypercube designs for the Kullback–Leibler criterion. *AStA Advances in Statistical Analysis* **94**(4), 341–351 (2010). DOI 10.1007/s10182-010-0145-y
- [57] Kai-Tai, F.: Theory, Method and Application of the Uniform Design. *International Journal of Reliability, Quality and Safety Engineering* **09**(04), 305–315 (2002). DOI 10.1142/s0218539302000858
- [58] Kebisek, M., Tanuska, P., Spendla, L., Kotianova, J., Strelec, P.: Artificial Intelligence Platform Proposal for Paint Structure Quality Prediction within the Industry 4.0 Concept. *IFAC-PapersOnLine* **53**(2), 11168–11174 (2020). DOI 10.1016/j.ifacol.2020.12.299
- [59] Kinnebrock, W.: Optimierung mit genetischen und selektiven Algorithmen. De Gruyter (1994). DOI 10.1515/9783486785203
- [60] Kleijnen, J.P.C.: Design and Analysis of Simulation Experiments. Springer-Verlag GmbH (2015)
- [61] Kleijnen, J.P.C., Mehdad, E.: Estimating the variance of the predictor in stochastic Kriging. *Simulation Modelling Practice and Theory* **66**, 166–173 (2016). DOI 10.1016/j.simpat.2016.03.008

- [62] Kleppmann, W.: *Versuchsplanung*. Carl Hanser Verlag GmbH & Co. KG (2020). DOI 10.3139/9783446463974
- [63] Kopper, A.E., Apelian, D.: Predicting Quality of Castings via Supervised Learning Method. *International Journal of Metalcasting* **16**(1), 93–105 (2021). DOI 10.1007/s40962-021-00606-7
- [64] Kumar, A.: Encoding schemes in genetic algorithm. *International Journal of Advanced Research in IT and Engineering* **2**(3), 1–7 (2013)
- [65] Kusne, A.G., Yu, H., Wu, C., Zhang, H., Hattrick-Simpers, J., DeCost, B., Sarker, S., Oses, C., Toher, C., Curtarolo, S., Davydov, A.V., Agarwal, R., Bendersky, L.A., Li, M., Mehta, A., Takeuchi, I.: On-the-fly closed-loop materials discovery via Bayesian active learning. *Nature Communications* **11**(1) (2020). DOI 10.1038/s41467-020-19597-w
- [66] Lavinias, Y., Aranha, C., Sakurai, T., Ladeira, M.: Experimental Analysis of the Tournament Size on Genetic Algorithms. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE (2018). DOI 10.1109/smc.2018.00617
- [67] Lekivetz, R., Jones, B.: Fast Flexible Space-Filling Designs for Nonrectangular Regions. *Quality and Reliability Engineering International* **31**(5), 829–837 (2014). DOI 10.1002/qre.1640
- [68] Liang, Y., Fang, K., Xu, Q.: Uniform design and its applications in chemistry and chemical engineering. *Chemometrics and Intelligent Laboratory Systems* **58**(1), 43–57 (2001). DOI 10.1016/s0169-7439(01)00139-3
- [69] Liu, S., Maljovec, D., Wang, B., Bremer, P.T., Pascucci, V.: Visualizing High-Dimensional Data: Advances in the Past Decade. *IEEE Transactions on Visualization and Computer Graphics* **23**(3), 1249–1268 (2017). DOI 10.1109/tvcg.2016.2640960
- [70] Lowe, D.: Radial basis function networks-revisited. *Mathematics Today* **51**(3), 124–126 (2015)
- [71] Lu, L., Anderson-Cook, C.M.: Strategies for sequential design of experiments and augmentation. *Quality and Reliability Engineering International* (2020). DOI 10.1002/qre.2823

- [72] Lu, L., Anderson-Cook, C.M., Martin, M., Ahmed, T.: Practical choices for space-filling designs. *Quality and Reliability Engineering International* (2021). DOI 10.1002/qre.2884
- [73] Lundkvist, P., Bergquist, B., Vanhatalo, E.: Statistical methods – still ignored? The testimony of Swedish alumni. *Total Quality Management & Business Excellence* **31**(3-4), 245–262 (2018). DOI 10.1080/14783363.2018.1426449
- [74] Medhi, T., Hussain, S.A.I., Roy, B.S., Saha, S.C.: Selection of best process parameters for friction stir welded dissimilar Al-Cu alloy: a novel MCDM amalgamated MORSM approach. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* **42**(10) (2020). DOI 10.1007/s40430-020-02631-9
- [75] Michiels, W., Aarts, E., Korst, J.: *Theoretical Aspects of Local Search*. Springer Berlin Heidelberg (2007). DOI 10.1007/978-3-540-35854-1
- [76] Montepiedra, G.: Application of genetic algorithms to the construction of exact D-optimal designs. *Journal of Applied Statistics* **25**(6), 817–826 (1998). DOI 10.1080/02664769822800
- [77] Montgomery, D.: *Design and analysis of experiments*. John Wiley & Sons, Inc (2017)
- [78] Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B.: Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences* **116**(44), 22071–22080 (2019). DOI 10.1073/pnas.1900654116
- [79] Nelles, O.: LOLIMOT-Lokale, lineare Modelle zur Identifikation nichtlinearer, dynamischer Systeme. *at-Automatisierungstechnik* **45**(4), 163–174 (1997)
- [80] Nelles, O.: *Nonlinear System Identification: From Classical Approaches to Neural Networks, Fuzzy Models, and Gaussian Processes*. Springer Nature (2020)
- [81] Nguyen, N.K., Miller, A.J.: A review of some exchange algorithms for constructing discrete D-optimal designs. *Computational Statistics & Data Analysis* **14**(4), 489–498 (1992). DOI 10.1016/0167-9473(92)90064-m
- [82] Pabst, T.: *Datenbasierte Modellierung und Optimierung eines Bitumenofens für Laugenbehälter auf Basis von geplanten Versuchen*. Masterarbeit, FH Bielefeld (2021)

- [83] Pasolli, E., Melgani, F.: Gaussian process regression within an active learning scheme. In: 2011 IEEE International Geoscience and Remote Sensing Symposium. IEEE (2011). DOI 10.1109/igarss.2011.6049994
- [84] Patterson, M., Singh, P., Cho, H.: The current state of the industrial energy assessment and its impacts on the manufacturing industry. *Energy Reports* **8**, 7297–7311 (2022). DOI 10.1016/j.egy.2022.05.242
- [85] Petelet, M., Iooss, B., Asserin, O., Loredo, A.: Latin hypercube sampling with inequality constraints. *AStA Advances in Statistical Analysis* **94**(4), 325–339 (2010). DOI 10.1007/s10182-010-0144-z
- [86] Pietraszek, J., Radek, N., Goroshko, A.V.: Challenges for the DOE methodology related to the introduction of Industry 4.0. *Production Engineering Archives* **26**(4), 190–194 (2020). DOI 10.30657/pea.2020.26.33
- [87] Pronzato, L., Müller, W.G.: Design of computer experiments: space filling and beyond. *Statistics and Computing* **22**(3), 681–701 (2012)
- [88] Rasmussen, C.: Gaussian processes for machine learning. MIT Press (2006)
- [89] Rehr, J., Kirchengast, M., Celikovic, S., Sacher, S., Kruisz, J., Khinast, J., Horn, M.: Improving Pellet Quality in a Pharmaceutical Hot Melt Extrusion Process via PID Control and LOLIMOT-Based MPC. *Journal of Pharmaceutical Innovation* **15**(4), 678–689 (2019). DOI 10.1007/s12247-019-09417-0
- [90] Rennen, G.: Subset selection from large datasets for Kriging modeling. *Structural and Multidisciplinary Optimization* **38**(6), 545–569 (2008). DOI 10.1007/s00158-008-0306-8
- [91] Santner, T.J., Williams, B.J., Notz, W.I.: The Design and Analysis of Computer Experiments. Springer New York (2018). DOI 10.1007/978-1-4939-8847-1
- [92] Scholz, D.: Optimierung interaktiv. Springer (2018)
- [93] Schreiter, J., Nguyen-Tuong, D., Eberts, M., Bischoff, B., Markert, H., Toussaint, M.: Safe exploration for active learning with Gaussian processes. In: Joint European conference on machine learning and knowledge discovery in databases, pp. 133–149. Springer (2015)

- [94] Schuster, E.F.: Incorporating support constraints into nonparametric estimators of densities. *Communications in Statistics - Theory and Methods* **14**(5), 1123–1136 (1985). DOI 10.1080/03610928508828965
- [95] Scott, D.W.: *Multivariate Density Estimation*. Wiley (2015)
- [96] Sequenz, H.: Emission Modelling and Model-Based Optimisation of the Engine Control, *VDI Fortschrittsberichte*, vol. 8. VDI Verlag, Düsseldorf (2013)
- [97] Sexton, C.J., Anthony, D.K., Lewis, S.M., Please, C.P., Keane, A.J.: Design of Experiment Algorithms for Assembled Products. *Journal of Quality Technology* **38**(4), 298–308 (2006). DOI 10.1080/00224065.2006.11918619
- [98] Sexton, C.J., Lewis, S.M., Please, C.P.: Experiments for derived factors with application to hydraulic gear pumps. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **50**(2), 155–170 (2001). DOI 10.1111/1467-9876.00226
- [99] Shang, C., You, F.: *Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era*. Engineering (2019)
- [100] Sharma, P., Chhillar, A., Said, Z., Memon, S.: Exploring the Exhaust Emission and Efficiency of Algal Biodiesel Powered Compression Ignition Engine: Application of Box–Behnken and Desirability Based Multi-Objective Response Surface Methodology. *Energies* **14**(18), 5968 (2021). DOI 10.3390/en14185968
- [101] Shina, S.: *Industrial Design of Experiments*. Springer International Publishing (2022). DOI 10.1007/978-3-030-86267-1
- [102] Shinde, G.V., Arakerimath, R.R.: Multi-response optimization of friction stir welding process of dissimilar AA3003-H12 and C12200-H01 alloys using full factorial method. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering* **235**(5), 1555–1564 (2021). DOI 10.1177/09544089211008280
- [103] Siebertz, K., van Bebber, D., Hochkirchen, T.: *Statistische Versuchsplanung*. Springer Berlin Heidelberg, Berlin and Heidelberg (2017). DOI 10.1007/978-3-662-55743-3
- [104] Silverman, B.W.: *Density estimation for statistics and data analysis*, vol. 26. CRC press (1986)

- 
- [105] Smits, V., Schussler, M., Kampmann, G., Illg, C., Decker, T., Nelles, O.: Excitation Signal Design and Modeling Benchmark of NO<sub>x</sub> Emissions of a Diesel Engine. In: 2022 IEEE Conference on Control Technology and Applications (CCTA). IEEE (2022). DOI 10.1109/ccta49430.2022.9966121
- [106] Sobol, I.M.: On quasi-Monte Carlo integrations. *Mathematics and Computers in Simulation* **47**(2-5), 103–112 (1998). DOI 10.1016/s0378-4754(98)00096-2
- [107] Som, A.M., Ramlee, A.A., Puasa, S.W., Hamid, H.A.A.: Optimisation of operating conditions during coagulation-flocculation process in industrial wastewater treatment using *Hylocereus undatus* foliage through response surface methodology. *Environmental Science and Pollution Research* (2021). DOI 10.1007/s11356-021-17633-w
- [108] Sturm, C.: Between a rock and a hard place: European energy policy and complexity in the wake of the Ukraine war. *Journal of Industrial and Business Economics* **49**(4), 835–878 (2022). DOI 10.1007/s40812-022-00233-1
- [109] Tharwat, A., Schenck, W.: Balancing Exploration and Exploitation: A novel active learner for imbalanced data. *Knowledge-Based Systems* **210**, 106500 (2020). DOI 10.1016/j.knosys.2020.106500
- [110] Thompson, K.R.: Implementation of gaussian process models for non-linear system identification. Ph.D. thesis, University of Glasgow (2009)
- [111] Viana, F.A.C.: Things you wanted to know about the Latin hypercube design and were afraid to ask. In: 10th World Congress on Structural and Multidisciplinary Optimization, vol. 19. sn (2013)
- [112] Voigt, T., Kohlhase, M., Nelles, O.: Inkrementelle Modellbildung von statischen Prozessen auf Basis von Latin Hypercube Designs. In: Proceedings - 29. Workshop Computational Intelligence, pp. 267–288. KIT Scientific Publishing, Karlsruhe (2019)
- [113] Voigt, T., Kohlhase, M., Nelles, O.: Incremental Latin Hypercube Additive Design for LOLIMOT. In: 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2020), pp. 1602–1609. IEEE (2020)

- 
- [114] Voigt, T., Kohlhase, M., Nelles, O.: Incremental DoE and Modeling Methodology with Gaussian Process Regression: An Industrially Applicable Approach to Incorporate Expert Knowledge. *Mathematics* **9**(19) (2021). DOI 10.3390/math9192479
- [115] Voigt, T., Schöne, M., Kohlhase, M., Nelles, O., Kuhn, M.: Using design of experiments to support the commissioning of industrial assembly processes. In: *Intelligent Data Engineering and Automated Learning – IDEAL 2022*, pp. 379–390. Springer International Publishing (2022). DOI 10.1007/978-3-031-21753-1\_37
- [116] Voigt, T., Schöne, M., Kohlhase, M., Nelles, O., Kuhn, M.: Space-Filling Designs for Experiments with Assembled Products. In: *2021 3rd International Conference on Management Science and Industrial Engineering*, pp. 192–199. ACM (2021). DOI 10.1145/3460824.3460854
- [117] Wang, W., Haaland, B.: Controlling Sources of Inaccuracy in Stochastic Kriging. *Technometrics* **61**(3), 309–321 (2018). DOI 10.1080/00401706.2018.1514328
- [118] Weicker, K.: *Evolutionäre Algorithmen*. Springer Fachmedien Wiesbaden (2015). DOI 10.1007/978-3-658-09958-9