

Reihe 12

Verkehrstechnik/
Fahrzeugtechnik

Nr. 810

Dipl.-Ing. Marek Kowalczyk,
Darmstadt

Online-Methoden zur effizienten Vermessung von statischen und dynamischen Verbrennungsmotor- modellen

Berichte aus dem

Institut für
Automatisierungstechnik
und Mechatronik
der TU Darmstadt



Online-Methoden zur effizienten Vermessung von statischen und dynamischen Verbrennungsmotormodellen

Vom Fachbereich
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von

Dipl.-Ing. Marek Kowalczyk

geboren am 16. September 1983 in Peiskretscham

Referent: Prof. Dr.-Ing. Dr. h. c. Rolf Isermann
Korreferenten: Prof. Dr. techn. Christian Beidl
Prof. Dr.-Ing. Ulrich Konigorski

Tag der Einreichung: 19. Juni 2017
Tag der mündlichen Prüfung: 22. Januar 2018



D 17

Darmstädter Dissertationen

Fortschritt-Berichte VDI

Reihe 12

Verkehrstechnik/
Fahrzeugtechnik

Dipl.-Ing. Marek Kowalczyk,
Darmstadt

Nr. 810

Online-Methoden zur
effizienten Vermessung
von statischen
und dynamischen
Verbrennungsmotor-
modellen

Berichte aus dem

Institut für
Automatisierungstechnik
und Mechatronik
der TU Darmstadt



Kowalczyk, Marek

Online-Methoden zur effizienten Vermessung von statischen und dynamischen Verbrennungsmotormodellen

Fortschr.-Ber. VDI Reihe 12 Nr. 810. Düsseldorf: VDI Verlag 2018.

192 Seiten, 68 Bilder, 3 Tabellen.

ISBN 978-3-18-381012-3, ISSN 0178-9449,

€ 67,00/VDI-Mitgliederpreis € 60,30.

Für die Dokumentation: Vermessung – Verbrennungsmotoren – Statisch – Dynamisch – Online-Methoden – Versuchsplanung – Aktives Lernen – Konkaver Variationsraum – Modellbildung – Optimierung

Die vorliegende Arbeit behandelt neue Methoden zur Reduktion des Modellierungsaufwandes und der Vermessungszeit an Motorenprüfständen. Hierfür wird eine iterative, aktiv lernende Methodik entwickelt, die erlaubt, sowohl das stationäre als auch dynamische Verhalten mit lokalen und globalen Modellen zu beschreiben. Mithilfe einer daten- und modellbasierten Messplanadaption unter Berücksichtigung von konkaven lokalen und globalen Variationsräumen sowie optimierten Stellgrößen wird der Aufwand und Umfang reduziert. Aus den daraus gemessenen Daten werden iterativ Modelle identifiziert, welche in relevanten Bereichen eine hohe Güte haben. Um die Anwendbarkeit der Methodik zu erhöhen und die benötigten Benutzereingriffe zu reduzieren, werden zusätzlich Methoden zur Anpassung der Modellstruktur entwickelt. Dadurch kann die erreichbare Modellgüte erhöht und Annahmen zur Vorausbestimmung der Modellstruktur vermieden werden. Die Anwendung der Online-Methodik wird exemplarisch für verschiedene statische und dynamische Beispiele vorgestellt.

Bibliographische Information der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter www.dnb.de abrufbar.

Bibliographic information published by the Deutsche Bibliothek

(German National Library)

The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at www.dnb.de.

D 17

© VDI Verlag GmbH · Düsseldorf 2018

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe (Fotokopie, Mikrokopie), der Speicherung in Datenverarbeitungsanlagen, im Internet und das der Übersetzung, vorbehalten.

Als Manuskript gedruckt. Printed in Germany.

ISSN 0178-9449

ISBN 978-3-18-381012-3

Danksagung

Die vorliegende Dissertation entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter bei Herrn Prof. Dr.-Ing. Dr. h.c. Rolf Isermann in der Forschungsgruppe Regelungstechnik und Prozessautomatisierung am Institut für Automatisierungstechnik und Mechatronik der Technischen Universität Darmstadt.

Mein Dank geht zuallererst an meinen Doktorvater Prof. Dr.-Ing. Dr. h.c. Rolf Isermann, der mir Vertrauen entgegengebracht und die Möglichkeit zur Promotion gegeben hat. Seine Unterstützung und Förderung, bei gleichzeitig eigenverantwortlicher Führungsweise, haben diese Arbeit ermöglicht und mich für meinen weiteren Berufsweg geprägt. Herrn Prof. Dr. techn. Christian Beidl, Leiter des Instituts für Verbrennungskraftmaschinen und Fahrzeugantriebe an der Technischen Universität Darmstadt, möchte ich für das Interesse an meiner Arbeit und die freundlichen Übernahme des Korreferats danken.

Weiterhin danke ich dem Bundesministeriums für Wirtschaft und Energie (BMWi), welches im Rahmen des Programms zur Förderung der Industriellen Gemeinschaftsforschung (IGF) über die Arbeitsgemeinschaft industrieller Forschungsvereinigungen (AiF) und die Forschungsvereinigung Verbrennungskraftmaschinen e.V. (FVV) mein Forschungsvorhaben finanziell unterstützt hat. An dieser Stelle gilt der Dank auch an den gesamten Arbeitskreis sowie insbesondere an den Obmann Herrn Dipl.-Ing. S. Lindner, die durch Ihren Blick auf die Themen aus der Perspektive der Industrie meine Arbeit geprägt haben.

Einen immensen Anteil an der Entstehung dieser Arbeit haben auch die vielen ehemaligen Kollegen am Institut. Dank der hilfsbereiten, unkomplizierten und freundschaftlichen Zusammenarbeit war der Alltag am Institut nie eintönig. Dies gilt sowohl für die VKM-Kollegen (Danke Florian, Simon, Andreas, Christopher, Matthias, Heiko, Alex), die sich für keine inhaltliche Diskussion oder Stunde am Prüfstand zu schade waren, als auch für die Nicht-VKM-Kollegen (Danke Markus, Mark, Carlo, Jacob, Philipp), die für den nötigen Blick über den Tellerrand gesorgt haben. Nicht zu vergessen natürlich die Damen (Danke Ilse, Brigitte, Corina, Sandra) und technischen Mitarbeiter (Danke Alfred, Alex), dank denen das Leben am Institut reibungslos ablief.

Abschließen möchte ich mit einem großen Dank an meine Familie. Sowohl an meine Eltern und an meinen Bruder, die mich stetig unterstützt und gefördert haben während des Studiums und der Promotion. Als auch an Julia und unseren Sohn Nils, die lange auf viel Verzicht mussten damit diese Arbeit zu Ende geführt werden konnte.

Darmstadt, Juni 2017

Marek Kowalczyk

Für Julia und meine Familie

Inhaltsverzeichnis

Symbole und Abkürzungen	VIII
Kurzfassung	XII
1 Einleitung	1
1.1 Stand der Technik	4
1.1.1 Elektronisches Motormanagement	4
1.1.2 Steuergerät-Applikation	5
1.1.3 Motorvermessung	6
1.1.4 Modellbasierte Sollwert-Optimierung	8
1.2 Zielsetzung und Aufbau der Arbeit	8
2 Gesamtstrategie zur Online-Vermessung	12
2.1 Prüfstandsanforderungen	13
2.2 Iterativer Ablauf der Gesamtstrategie	14
2.3 Schnittstellen und Struktur der Gesamtstrategie	16
2.3.1 Online-Vermessungssystem <i>ORTEM</i>	18
2.3.2 Online-Auswertungssystem <i>OTOM</i>	21
2.4 Zusammenfassung	23
3 Methoden zur Bestimmung zulässiger Aktor-Stellbereiche	24
3.1 Grundlagen der Variationsraumbestimmung	25
3.2 Grenzgrößen bei der Variationsraumbestimmung	27
3.2.1 Arten der Grenzgrößen	27
3.2.2 Bestimmung von Zündaussetzern als Grenzgröße	27
3.3 Messpläne zur Bestimmung von Variationsräumen	31
3.4 Schnelle Bestimmung des statischen Variationsraums	32
3.5 Ansätze zur einfachen Beschreibung von statischen konkaven Variationsräumen	37
3.5.1 Einfache konkave Variationsraumdarstellung	38
3.5.2 Dimensionsreduktion von Variationsräumen	42
3.5.3 Interpolation von Variationsräumen	43
3.6 Zusammenfassung	46
4 Online-Versuchsplanung bei der Motorvermessung	48
4.1 Grundlagen der statistischen Versuchsplanung und modellbasierten Optimierung	49
4.1.1 Arten der Datenaufzeichnung	49
4.1.2 Einflussgrößenauswahl und Effektermittlung	51
4.1.3 Messpunktverteilung	52
4.1.4 Anregungssignale zur Vermessung	56
4.1.5 Verfahren zur modellbasierten Optimierung	59

4.2	Anregungssignal zur Identifikation von stationären und dynamischen Modellen	60
4.2.1	Vergleich der Signale zur dynamischen Prozessanregung	63
4.3	Aktiv lernende Messpläne	65
4.3.1	Gesamtstrategie zur Online-Messplanerweiterung	66
4.3.2	Iterative Stellbereichsanpassung	67
4.3.3	Modellunabhängige Bestimmung der Messpunkte	72
4.3.4	Modellbasierte Bestimmung der Messpunkte	74
4.3.5	Zielgerichtete Messplanerweiterung im Rahmen der Online-Methodik	80
4.3.6	Beispiel der Online-Versuchsplanung	81
4.4	Zusammenfassung	83
5	Online-Identifikation von mathematischen Verbrennungsmotor-Modellen	85
5.1	Grundlagen der experimentellen Modellbildung	86
5.1.1	Parameterschätzung	87
5.1.2	Modellbildungsverfahren für stationäre Prozesse	89
5.1.3	Modellbildungsverfahren für dynamische Prozesse	95
5.1.4	Verfahren zur Bewertung der Modellgüte	98
5.1.5	Bias-Varianz Dilemma	100
5.1.6	Strukturbestimmung von Modellen	101
5.2	Online-Bias-Varianz-Dilemma	104
5.3	Online-Modellanalyse	105
5.4	Online lokal polynomiale Netzmodelle	107
5.4.1	Online-Adaption der Modellparameter	107
5.4.2	Online-Adaption der Regressoren	108
5.4.3	Online-Adaption der Netzmodellstruktur	108
5.4.4	Online-Adaption der Ordnung der Modelleingangsgroßen	114
5.4.5	Gesamtstrategie zur Online-Modellkomplexitätsadaption	114
5.4.6	Beispiel der Online-Identifikation	116
5.5	Zusammenfassung	119
6	Anwendungsbeispiel der Online-Vermessung	121
6.1	Parametrierung der Online-Vermessung	122
6.2	Verlauf der Online-Vermessung	123
6.3	Auswertung der Online-Vermessung	124
6.4	Vergleich der Online-Vermessung mit einer klassischen Offline-Vermessung	130
6.5	Zusammenfassung	132
7	Zusammenfassung und Ausblick	133
A	Motorenprüfstand und Versuchsträger	136
B	Software Dokumentation	138
C	Höherdimensionale konkave Variationsräume	141

D	Dynamische Variationsräume	144
E	Bestimmung der Abtastzeit ohne Prozesswissen bei dynamischen Messungen	146
F	Datenvorbereitung	149
G	Erweiterungen der LS-Parameterschätzung	150
H	Vermessung von Kennfeldmodellen	152
I	Identifikation von richtungsabhängigen Prozessen	157
	Literaturverzeichnis	161

Symbole und Abkürzungen

Symbole

A	Anzahl an Ausgangsgrößen in einem System/Modell
c	Kennfeldstützstellenposition, Geometrischer Schwerpunkt eines Simplex
C	Anzahl an verzögerten Werten der Eingangsgröße
D	Dynamische Ordnung eines Polynoms Anzahl an verzögerten Werten der Ausgangsgröße
e_i	Residuum
E	Anzahl an Eingangsgrößen in einem Modell
$G_{\text{RMSE,global}}$	Grenzwert des globalen $J_{\text{RMSE,global}}$
H	Teilmenge an Stellgrößen in einem System
i, j, o, p	Zählvariable
\mathbf{I}	Einheitsmatrix
J_A	Kenngroße der Adaption
J_{MAE}	Mittlerer absoluter Fehler
$J_{\text{RMSE,lokal}}$	Wurzel des mittleren quadratischen lokalen Fehlers
$J_{\text{RMSE,global}}$	Wurzel des mittleren quadratischen globalen Fehlers
J_Q	Kenngroße der Modellgüte
J_S	Kenngroße der Messgrößenstandardabweichung
k	Diskrete Zeit $k = t/T_0 = 0, 1, 2, \dots$
\mathbf{K}	Kandidatensatz mit potentiellen Stellgrößenkombinationen
l	Teilungsverhältnis, Streckfaktor
L	Stufen eines Rasters Anzahl an Teildatensätzen Anzahl an lokalen Teilmodellen
M	Anzahl an lokalen Teilmodellen
\mathbf{M}	Messplan, bestehend aus zu vermessenden Stellgrößenkombinationen
\mathbf{M}_0	Initialer Messplan zu Beginn der Vermessung
n	Stufenanzahl eines Schieberegisters Zählvariable
n_{Mot}	Drehzahl des Verbrennungsmotors in min^{-1}
N	Anzahl an Zeitintervallen beim APRBS Anzahl an gesamten Stellgrößenkombinationen
O	Anzahl an ausgewählten bzw. zusätzlichen Stellgrößenkombinationen
P	Anzahl an Stellgrößen in einem System Leistung in kWh
p_{Zyl}	Zylinderdruck
$\Delta p_{\text{Zyl,max}}$	Maximale Differenz des Zylinderdrucks
q_{HE}	Haupteinspritzmenge in $\frac{\text{mm}^3}{\text{Hub}}$
Q_s	Ladung im piezoelektrischen Quarzsensor

Q	Anzahl an gesamten Regressoren
r_o	Radius der o -ten Stellgrößenkombination in Kugelkoordinaten
R	Anzahl ausgewählter Regressoren
R^2	Bestimmtheitsmaß
S	Statische Ordnung eines Polynoms
$s_{\text{Aussetzer}}$	Anzahl an auftretenden Aussetzern
t	Kontinuierliche Zeit in s
T_0	Abtastzeit in s
T_E	Einschwingzeit in s
T_p	Periodenlänge in s
u	Stellgröße eines Systems
\mathbf{u}_+	Eingangsgrößenvektor aus erweitertem Datensatz
$v_{\text{Aussetzer}}$	Auftretender Aussetzern
V	Anzahl an gesamten Kennfeldstützstellen
w	Kennfeldstützstellenhöhe
\mathbf{W}	Gewichtungsmatrix
x	Eingangsgröße eines Modells
y	Gemessene Ausgangsgröße eines Systems
y_{sim}	Simulierte Ausgangsgröße eines Systems
\bar{y}	Gemittelte Ausgangsgröße eines Systems
\hat{y}	Ausgangsgröße eines Modells
\hat{y}_{QbC}	Ausgangsgröße eines Modellkomitees
\mathbf{y}_+	Ausgangsgrößenvektor aus erweitertem Datensatz
Z	Anzahl an Iterationen
$\hat{\theta}$	Parameter eines Modells
$\hat{\theta}_+$	Parameter eines Modells nach der Schätzung mit erweitertem Datensatz
λ	Regularisierungsparameter
λ_{\min}	Taktzeit eines APRBS
λ_{\max}	Maximale Haltezeit eines APRBS
μ	Arithmetrischer Mittelwert
σ	Standardabweichung
σ^2	Varianz
σ_{seg}^2	Varianz in einem Segment
σ_{lim}^2	Schwelle der zulässigen Varianz
φ_o	Winkel der o -ten Stellgrößenkombination in Kugelkoordinaten
φ_{KW}	Winkel der Kurbelwelle
φ_{OT}	Winkel der Kurbelwelle bei oberem Totpunkt
φ_{VB}	Winkel der Kurbelwelle bei Verbrennungsbeginn
Ψ	Datenmatrix zur Parameterschätzung
Ψ_+	Datenmatrix aus erweitertem Datensatz zur Parameterschätzung

Abkürzungen

ADBS	Amplitudenmoduliertes Deterministisches Binär-Signal
ADRBS	Amplitudentmoduliertes Diskretes-Rausch-Binär-Signal
AGRBS	Amplitudentmoduliertes Generalisiertes-Rausch-Binär-Signal
AIC	Akaikes' information criterion
APRBS	Amplitudentmoduliertes Pseudo-Rausch-Binär-Signal
DBS	Deterministisches Binär-Signal
DoE	Versuchsplanung (design of experiments)
DTS	Delaunay triangulation sampling
GMR	Globale Modellregion
HHT	Hinging hyperplane trees
LASSO	least absolute shrinkage and selection operator
LARS	least angle regression
LHS	Latin hypercube sampling
LLM	Lokal lineares Modell
LM	Lokales Modell
LMR	Lokale Modellregion
LOLIMOT	Lokal lineares Netzmodell (local linear model tree)
LOPOMOT	Lokal polynomiales Netzmodell (local polynomial model tree)
LPM	Lokal polynomiales Modell (local polynomial model)
LS	Methode der kleinsten Quadrate (least squares)
MAE	Mean absolute error
MIMO	Mehrere Ein- und Ausgangsgrößen (multi input multiple output)
MISO	Mehrere Eingangs-, eine Ausgangsgröße (multi input single output)
MLP	Multilayer-Perzeptron-Netze
MO	Model optima
MPC	Model parameter change
MQD	Model quality decrease
MQI	Model quality increase
MSE	Mean squared error
NEFZ	Neuer Europäischer Fahrzyklus
NMSE	Normalized mean squared error
NO _x	Stickstoffoxide (NO und NO ₂)
NRMSE	Normierte Wurzel des mittleren quadratischen Fehlers
NSGA	Non-dominated sorting genetic algorithm
ONLOPOMOT	Online lokal polynomiales Netzmodell (o. local polynomial model tree)
ONPOLY	Online Polynommodell (online polynomial model)
OTOM	Online target oriented measurement
ORTEM	Online real-time engine measurement
PRBS	Pseudo-Rausch-Binär-Signal
PRESS	Predicted residual error sum of squares
PRMS	Pseudo-Rausch-Mehrstufen-Signal

RBF	Radial-Basis-Funktions-Netze
RCP	Rapid Control Prototyping
RLS	Rekursive Methode der kleinsten Quadrate
RMSE	Wurzel des mittleren quadratischen Fehlers (root mean squared error)
SAE	Summed absolute error
SIMO	Eine Eingangs-, mehrere Ausgangsgrößen (single input multiple output)
SISO	Eine Ein- und Ausgangsgröße (single input single output)
SNR	Signal-Rausch-Verhältnis (signal noise ratio)
VR	Variationsraum

Kurzfassung

Die Reduktion von Abgas-Emissionen sowie von Kraftstoffverbrauch gewinnt immer stärkere Bedeutung und führt in der Umsetzung zu einer immer höheren Anzahl an Aktoren und Sensoren und infolgedessen zu einer immensen Zunahme an Variabilitäten. Die Applikation von elektronischen Motorsteuerungen wird folglich zu einer hochdimensionalen Aufgabe und mit der Dimensionalität steigt die Komplexität. Um die Optimierung der Applikation in annehmbarer Zeit und mit akzeptablen Aufwand durchführen zu können sind sowohl schnelle Vermessungsmethoden als auch modellbasierte Optimierungsmethoden nötig. Die Zeit und der Aufwand können weiter reduziert werden, indem die Messung, Identifikation und Optimierung zu einer Online-Methodik am Motorenprüfstand kombiniert werden, wodurch diese automatisch interagieren und sich anpassen.

Zur Umsetzung einer schnellen Applikation wurde eine Online-Methode entwickelt, welche die Modellierung des statischen und dynamischen Verhaltens mit lokalen sowie globalen Verbrennungsmotormodellen ermöglicht. Die Automatisierung erfolgt dabei durch die Parallelisierung der Vermessung und Auswertung zur Modellbildung, sodass sie ohne Trennung von Messungen am Prüfstand, Auswertungen im Büro und zwischenzeitlichen Benutzereingriffen ablaufen kann. Der Ablauf des Prozesses entspricht einem iterativen Vorgehen während der Vermessung, wodurch passende Offline-Methoden verwendet werden können. Mithilfe verschiedener weiterer Methoden kann eine Reduktion des gesamten Aufwands und der benötigten Zeit erreicht werden, bei gleichzeitiger Erhöhung der Modellqualität.

Eine schnelle Bestimmung des Variationsraums wird durch eine limitbasierte Regelung der Aktorführungsgrößen erreicht. Solange die gemessenen Größen weit von ihren kritischen Werten entfernt sind, wird die Verstellgeschwindigkeit aller Aktoren erhöht, um schneller zu den Variationsraumgrenzen zu gelangen. Wenn die gemessenen Größen sich ihren kritischen Werten nähern, wird die Verstellgeschwindigkeit geregelt reduziert. Mithilfe dieser integrierten Regelung wird ein schnelles Verfahren erreicht. Mithilfe der gemessenen Positionen der Aktoren an den Variationsraumgrenzen wurde eine mathematische Beschreibung des konkaven Variationsraums und der sich daraus ergebenden konkaven Hülle entwickelt. Speziell bei hochdimensionalen Problemen ist die mathematische Beschreibung des konkaven Variationsraums essentiell für eine automatische Online-Vermessung am Prüfstand, weil ein manuelles Vermessen hochdimensionaler Räume viel Zeit in Anspruch nimmt. Mithilfe einer Stationärwerterkennung kann die stationäre Vermessung schneller erfolgen. Die Online-Methodik erkennt automatisch den stationären Zustand aller Messgrößen, führt die Datenaufzeichnung aus und wechselt zum nächsten Messpunkt. Infolgedessen wird keine globale Stabilisierungszeit mehr benötigt, da die Messgrößen bis zum Erreichen aller stationären Werte für eine individuelle Zeit konstant gehalten werden. Für die dynamische Modellbildung wurde ein Kurzzeit-Testsignal entwickelt, welches eine flexible und zielorientierte Anpassung erlaubt. Ohne eine individuell angepasste Anregung können dynamische Modelle keine hohe Güte erreichen. Darüber hinaus wurden geeignete Methoden zur zielgerichteten sowie modellbasierten Anpassung von Messplänen basierend auf

dem aktuell gültigen Modell entwickelt. Auf diese Weise wird eine schnellere Vermessung zur Modellbildung ermöglicht, da statt klassischen raumfüllenden Designs nur Messpunkte erfasst werden, welche für das Modell und die Anwendung relevant sind. Der gesamte Messaufwand kann reduziert werden, indem das Messen in Bereichen hoher Modellgüte vermieden wird, da der Einfluss einer neuen Messung in diesen Bereich gering ist. Um die Anwendbarkeit der Methodik zu erhöhen sowie die benötigten Benutzereingriffe zu reduzieren, werden zusätzlich Methoden zur Anpassung der mathematischen Modellstruktur entwickelt. Dadurch kann die erreichbare Modellgüte erhöht werden und Annahmen zur Modellstruktur werden vermieden.

Mit der hier vorgestellten neuen Methodik zur zielgerichteten Vermessung, welche sowohl schnell als auch automatisch abläuft, leistet diese Dissertation einen wesentlichen Beitrag zur modellbasierten Applikation von Motorsteuergeräten. Durch den Einsatz dieser Online-Methodik, deren Anwendbarkeit anhand verschiedener Beispiele verdeutlicht wird, kann die Vermessung und Identifikation in deutlich kürzerer Zeit als beim klassischen Vorgehen erfolgen.

Abstract

The reduction of exhaust emissions and fuel consumption is gaining in importance and leads in the implementation to an increasing number of actuators and sensors and consequently to an immense increased number of variabilities. The calibration of the electronic engine control units becomes a high dimensional task and with the dimensionality, the complexity increases. To complete the optimization of the calibration in an acceptable amount of time and reasonable effort, fast measurement methods as well as model-based optimization methods are needed. Even more time and effort can be saved, if the measurement, identification and optimization are coupled to an online-method at an engine test bench, resulting in an automatic interaction and adaptation of them.

For enabling a fast calibration, an online methodology was developed, which allows the modeling of the stationary as well as the dynamic behaviour with local and global combustion engine models. The automatization is done by parallelization of the measuring and analysis for identification, enabling a procedure without separation of measuring at the test bench and evaluation at the office and user inputs in between. The sequence of the process corresponds to an iterative approach during the measurement, which enables the use of suitable offline methods. With the help of different additional methods a reduction of the overall effort and time is achieved, while enhancing the model quality.

A fast determination of the variation space is accomplished with a limit based control of actuators. As long as the measured signals are far away from their critical values, the adjustment speed of all actuators is increased to move faster towards the variation space border. If the measured signals approach their critical values, a controlled reduction of the speed is performed. With the integrated controller, a fast determination of the variation space is accomplished. With the measured positions of the actuators at the variation space borders, a mathematical description of the concave variation space and hull have been developed. Especially for high dimensional problems, the mathematical description of the concave variation space is essential for an automated online-measurement at the test bench, as the manual measurement of high-dimensional spaces takes much time. With the help of a stationary detection, stationary measurements can be performed faster. The online methodology recognizes automatically the stationary value of all measured variables, performs the data acquisition and continues to the next measurement point. No global stabilization time is needed any more as the measured values are held for an individual time until they all are stable. For dynamic modeling, a short-time excitation signal is developed, which enables a flexible and target-oriented parametrization. Without an individually adapted excitation, dynamic models can't achieve a good quality. Furthermore, suitable methods for the target-oriented and model-based adaptation of the measurement design based on the actual valid model are developed. This enables a faster measurement for modeling, as instead of classic variationspace-filling designs only points are captured which are relevant for the actual model and application. A reduction of the measurement effort is achieved by avoiding regions where the quality of the model is already high, as the influence of new measurement

points in this area is small. For increasing the applicability of the methodology and reducing the necessary user inputs, methods for adaptation of the mathematical model structure are developed. This increases the achievable model quality and assumptions for the model structure are avoided.

With the here presented new measurement methodology for target-oriented measurements, which performs fast and fully automatically, this dissertation contributes essentially to the model-based calibration of engine control units. Using the online methodology, whose applicability is exemplarily shown with different examples, the measurement and modeling can be accomplished in significantly lesser time than with common approaches.

1 Einleitung

Diese Arbeit behandelt die Optimierung des Vermessungsprozesses zur Bestimmung von Verbrennungsmotormodellen, welche im Rahmen der Applikation und Kalibrierung von Motorsteuergeräten eingesetzt werden. In diesem Kapitel wird die Motivation zur Nutzung von modellbasierten Methoden dargelegt. Hierfür werden die Emissionsgesetzgebung sowie die sich daraus ergebenden Anforderungen an die Motorentwicklung diskutiert und das V-Modell der Funktionsentwicklung für Steuergeräte verdeutlicht. Daraus resultierend wird der Vorteil des Einsatzes von modellbasierten Methoden zur durchgängigen Vermessung von Motoren dargestellt. Es wird ein Überblick über den Stand der Technik der modellbasierten Steuergeräteentwicklung sowie über die Inhalte der vorliegenden Arbeit gegeben.

Die Entwicklung moderner Verbrennungsmotoren und Antriebskonzepte ist in den letzten Jahren einer zunehmenden Anzahl von Anforderungen unterworfen. Heutige Motorkonzepte sind durch einen hohen technologischen Stand und eine dementsprechend hohe Komplexität geprägt. Verantwortlich sind hierfür zum Einen die langfristig steigenden Kraftstoffpreise und die infolgedessen stärkere Bedeutung von Kraftstoffverbrauchs- und Emissionsreduktion. Zum Anderen trägt maßgeblich die sich stetig verschärfende Abgas- sowie Überwachungsgesetzgebung bei gleichbleibend guter bzw. oftmals verbesserter Fahrbarkeit dazu bei.

Bild 1.1 zeigt die Entwicklung der Europäischen Emissionsgesetzgebung für PKW-Dieselmotoren sowie die Preisentwicklung für Diesel- und Superkraftstoff in den letzten Jahren. Die Emissionsgesetzgebung in Europa wird von der Europäischen Kommission definiert. Dabei werden verschiedene Abgaskomponenten (Kohlenmonoxid CO, Stickstoffoxide NO_x, Kohlenwasserstoffe HC, Partikelmasse des Feinstaubes PM sowie seit 2011 die Partikelanzahl PN) über den Neuen Europäischen Fahrzyklus (NEFZ) limitiert. Die Klassifikation der Verbrennungsmotoren (EURO 1 in 1992 bis EURO 6 in 2014) erfolgt dabei abhängig davon, welche Grenzen (in kg/km) die Abgaskomponenten im Zyklus einhalten. Wie in Bild 1.1 ersichtlich, wurden in den letzten Jahren die Grenzen immer weiter verschärft. Speziell die Reduktion der Grenzwerte von NO_x und gleichzeitig PM, aufgrund der vorhandenen Trade-Offs der beiden Größen, stellt die Entwicklung neuer Motoren sowie derer Steuergeräte vor eine große Herausforderung, da eine Größe nicht verringert werden kann, ohne die Zweite zu erhöhen. Voraussichtlich im Jahr 2017 wird der NEFZ, dessen Leerlaufzeit 33% beträgt, durch den WLTC in der Klassifikation ersetzt, der nur noch 13% Leerlaufzeit hat und einen größeren dynamischen Anteil aufweist. Der neue Zyklus, zusammen mit den in der Diskussion befindlichen RDE (real world driving emissions), macht es zwingend erforderlich, neben dem stationären Verhalten das dynamische Verhalten bei der Entwicklung zu integrieren.

Die hohe Anzahl an zusätzlichen Komponenten führt zwangsläufig zu einem höheren Aufwand und einer größeren Komplexität in der Fehlerüberwachung und -diagnose. Dabei spielt zum

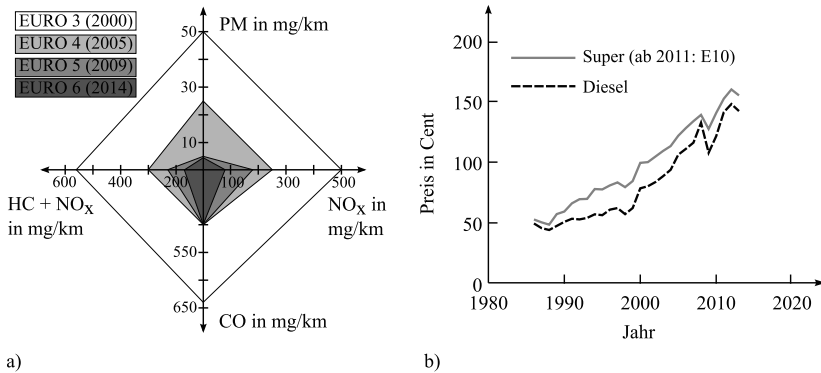


Bild 1.1: a) Entwicklung der Grenzwerte für die vom Gesetzgeber limitierten Dieselmotor-Emissionen seit dem Jahr 2000 [34, 35]. b) Entwicklung der Preise für Diesel und Super seit dem Jahr 1988 [1].

Einen die Überwachung der Emissionen im Fehlerfall eine Rolle sowie zum Zweiten die Möglichkeit, die Art, Größe und Position des Fehlers im komplexen Gesamtsystem lokalisieren zu können. Treibende Kraft der Entwicklung sind dabei nicht nur die Gesetzgeber, die seit der 90er Jahre in Form der on-board-Diagnose (OBD) die Überwachung aller abgasrelevanten Komponenten gesetzlich vorschreiben, sondern auch die Kunden, welche bei der Behebung von Fehlern nicht nur weniger Kosten, sondern auch weniger Zeit aufwenden wollen.

Die PKW-Hersteller haben zur Erfüllung der Anforderungen verschiedene Schritte unternommen. Neben der Optimierung von eingesetzten Werkstoffen (Alu) und der Entwicklung neuer Brennverfahren (z.B. HCCI) wurden vor allem verschiedene neue Technologien eingeführt. In der Gesamtheit hat diese Entwicklung dazu geführt, dass viele neue Sensoren und Aktoren im Motor vorhanden sind, wodurch die Anzahl an Variabilitäten exponentiell gestiegen ist. Die Steuerung und Überwachung der Variabilitäten wird dabei von einem digitalen Motorsteuergerät übernommen. Zur Erfüllung der Steuer-, Regel- und Diagnosekonzepte ist sowohl deren Leistungsfähigkeit als auch deren Speicherkapazität in den letzten Jahren exponentiell gestiegen (siehe Bild 1.2).

Entsprechend dem Anstieg der Applikationsparameter ist auch der Aufwand zur Generierung sowie Abstimmung aller Steuer-, Regel- und Diagnosefunktionalitäten im Steuergerät angestiegen. Die Generierung von Sollwert-Kennfeldern und der Entwurf sowie die Kalibrierung von Steuerungen und Regelungen stellen inzwischen eine nichtlineare Systemoptimierung mit vielen Freiheitsgraden dar. Diese hoch-dimensionale Aufgabe der Applikation sowie die optimale Einstellung aller Stellgrößen und Diagnosen in allen relevanten Betriebspunkten unter Einhaltung der Abgasgrenzwerte kann nicht mehr manuell am Prüfstand bestimmt werden. Dies würde ein nicht mehr vertretbaren zeitlichen Aufwand darstellen. Erschwert wird die Bestimmung durch weltweit unterschiedliche Abgasgesetzgebungen (z.B. CARB und EPA in Nordamerika oder Post New Long Term in Japan), die verschiedene Motorapplikationen erfordern.

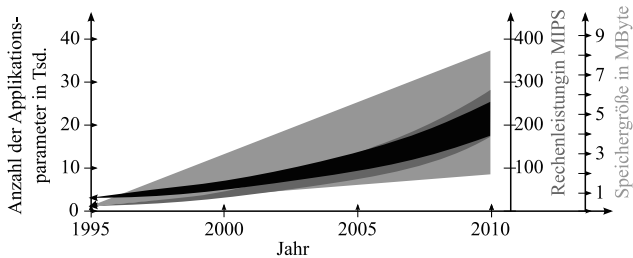


Bild 1.2: Entwicklung der Leistungsfähigkeit von Steuergeräten seit 1995. Sowohl die Anzahl der Applikationsparameter (Kennwerte von Funktionsalgorithmen) als auch die Rechenleistung steigen exponentiell, während die Speichergröße linear zunimmt [33].

Die Ausnutzung des großen Potentials der zusätzlichen Variabilitäten und Freiheitsgrade in modernen Motoren kann nur noch durch den Einsatz von modellbasierten Methoden im Entwurfs- und Applikationsprozess erfolgen. In Bild 1.3 ist das V-Modell für die Funktionsentwicklung von Steuergeräten dargestellt. Dieses zeigt, dass ausgehend von der Spezifikation ein Entwurf am PC, die Implementierung als C-Code sowie die Seriencodegenerierung für das Steuergerät erfolgen. Sobald die Funktion als Zielsystem-Code vorliegt, beginnt das Testen, die Kalibrierung sowie die Integration. Dabei ist die erneute Durchführung der Schritte vom Entwurf bis zur Implementierung notwendig, um alle Anforderungen und Spezifikationen zu erfüllen. Steht bereits beim Funktionsentwurf ein Modell des Verbrennungsmotors zur Verfügung, kann die Anzahl der benötigten Korrekturschleifen reduziert werden, da die korrekte Arbeitsweise der Funktion bereits mithilfe des Modells erreicht werden kann.

Der herkömmliche Entwicklungsprozess wird den verkürzten Entwicklungszyklen unter vermehrtem Kostendruck sowie immer komplexeren Systemen und anspruchsvolleren Aufgabenstellungen der Applikation nicht mehr ausreichend gerecht. Zudem erfordert die teure Prüfstandszeit neben dem Einsatz von modellbasierten Methoden weitere Methoden, die den Entwurf der Modelle zur Applikation und Kalibrierung verkürzen, ohne dabei die Applikationsgüte wesentlich zu verringern. Hierfür muss der Prozess zur Bestimmung der benötigten Motormodelle optimiert werden, indem dieser effizient gestaltet und die Prüfstandszeit effektiv ausgenutzt wird. Durch den Einsatz von automatisierten dynamischen Prüfständen, von mobiler dynamischer Abgasmesstechnik sowie von innovativen modellbasierten Methoden zur Vermessung und Modellbildung können, kombiniert mit optimierten Entwicklungsprozessen, kürzere Entwicklungszyklen erreicht werden (siehe Bild 1.4). Dadurch kann wiederum der zukünftige Entwicklungsaufwand, die benötigte Entwicklungszeit und die daraus resultierenden Entwicklungskosten klein gehalten werden.

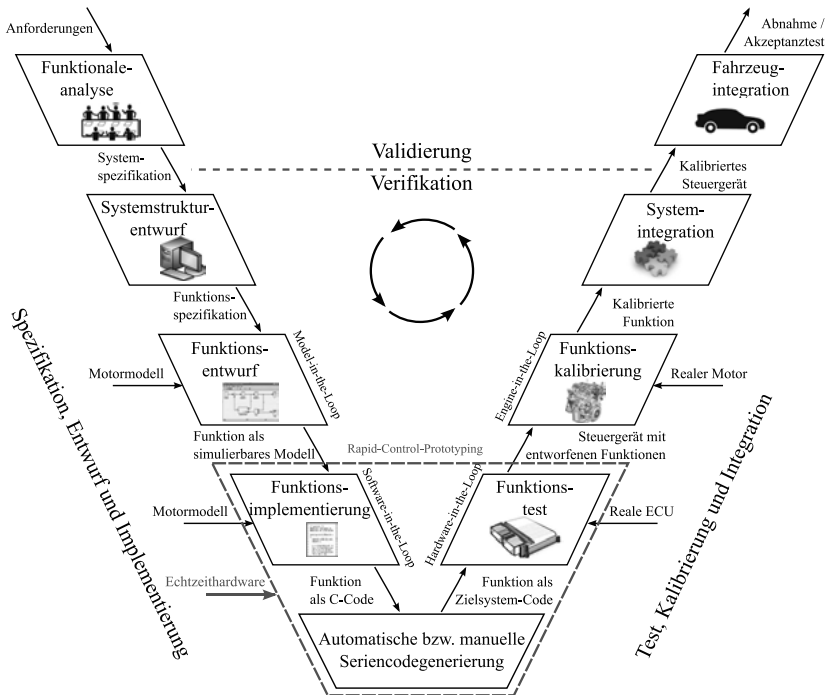


Bild 1.3: V-Modell zur Funktionsentwicklung von Verbrennungsmotorsteuergeräten

1.1 Stand der Technik

1.1.1 Elektronisches Motormanagement

Die digitale Motorsteuerung (ECU, Engine Control Unit) übernimmt die Steuerung, Regelung und Diagnose aller im Verbrennungsmotor vorhandenen Prozesse. Während 1979 digitale Steuergeräte nur in Oberklassefahrzeugen verbaut worden sind, können heutige Motoren aufgrund ihrer Komplexität und der gegebenen Anforderungen nicht mehr ohne digitale Steuergeräte betrieben werden. Sensoren erfassen dabei alle wichtigen, messbaren Zustandsgrößen (Drücke, Temperaturen, Positionen, Drehzahlen, Luftmassen) analog oder digital. Die Motorsteuerung verarbeitet und wertet die Signale über programmierte Funktionen aus. Dabei wird der aktuelle Ist-Zustand des Motors ermittelt, und basierend auf den Anforderungen des Fahrers und der Nebenaggregate werden die Sollwerte für die Stellsignale der Aktoren (Elektromechanische-, Piezo- oder Pneumatik-Aktoren) generiert. Die Aktoren stellen dann basierend auf den Vorgaben der Motorsteuerung alle einzelnen Motorkomponenten (z.B. Einspritzventil, Drosselklappe, AGR-Ventil, VTG-Leitschaufeln, Nockenwellensteller, etc.). In der Regel wird das gesamte Teilsystem, bestehend aus Stellmotor und entsprechender Mechanik, als Aktor bezeichnet. Die

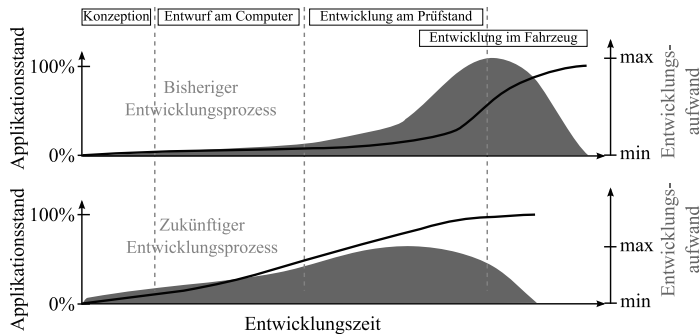


Bild 1.4: Bisheriger und zukünftiger Entwicklungsprozess. Durch eine Verschiebung in die frühere Phase kann die benötigte Zeit und der maximale Aufwand reduziert werden.

Sensorik und Aktorik sowie der Aufbau des Steuergeräts sind dabei stark vom Motorkonzept abhängig. Die Software der Motorsteuerung wird in einen festen Programmcode und frei veränderliche Parameter unterteilt. Die veränderlichen Parameter können dabei neben einzelnen Konstanten auch Kennlinien oder Kennfelder sein. Dadurch kann der vom Funktionsentwickler erstellte Programmcode für verschiedenen Motoren genutzt werden, während der Applikateur nur noch die veränderlichen Parameter je nach Motortyp und gewünschtem Motorverhalten einstellen muss. Das Vorgehen wird entsprechend als Funktionsentwicklung bzw. Applikation bezeichnet.

1.1.2 Steuergerät-Applikation

Die Steuergerät-Applikation erfolgt durch optimale Einstellung von Stellgrößen für einen Motor. Sie kann generell in 4 Schritte unterteilt werden:

1. Basisapplikation: Grundparametrierung der relevanten Motorsteuerungsfunktionen zu einem sicheren Betrieb mit geringem Verbrauch und Emissionen.
2. Fahrbarkeitsapplikation: Anpassung aller Komfortfunktionen und Funktionen, welche für mehrere Komponenten relevant sind (z.B. Schaltvorgänge für Getriebe und Motor).
3. Emissions- und Verbrauchsoptimierung: Optimierung für definierte Fahrzyklen.
4. Diagnoseapplikation: Parametrierung aller für die Diagnose notwendigen Parameter.

Im Fall von einfachen Problemstellungen, d.h. wenn der Zusammenhang zwischen den Funktionen, einstellbaren Parametern und dem Prozess analytisch lösbar ist, kann das Suchen und Finden der optimalen Einstellung direkt am Prüfstand erfolgen. Da der Zusammenhang zwischen dem Prozess und den einstellbaren Parametern in der Regel nicht analytisch lösbar ist,

muss eine iterative heuristische, auf Expertenwissen basierende, Vorgehensweise angewandt werden. In [78] wurde dieses Vorgehen für die stationäre Optimierung der Sollwerte direkt am Prüfstand gezeigt. Dabei wurde die Zielfunktion als Polynom zweiter Ordnung angenommen, ohne weitere Prozessannahmen, und diese direkt am Prüfstand minimiert. Dieses Vorgehen wurde als *online Optimierung* bezeichnet, da hierbei die Optimierung direkt am Prüfstand und nicht im Anschluss im Büro erfolgt. Nachteilig ist dabei die heuristische Wahl der Zielfunktion sowie die Tatsache, dass keine Information zur weiteren Nutzung in Form eines Modells vorliegt. Aufgrund der gestiegenen Komplexität hat sich in den letzten Jahren die modellbasierte Vorgehensweise durchgesetzt. Dabei wird basierend auf den Messdaten ein Motormodell identifiziert, mit Hilfe dessen die optimale Einstellung der freien Parameter mit rechnerbasierten Algorithmen erfolgen kann. Etabliert sind dabei Offline-Methoden, bei denen der Prüfstandsbetrieb und die rechnergestützte Datenauswertung getrennt erfolgen.

1.1.3 Motorvermessung

Zur modellbasierten Applikation wird ein Modell des Verbrennungsmotors benötigt. Je nach Anwendung des Modells und Komplexität des Motors werden verschiedene Modellstrukturen verwendet, deren Parameter entweder theoretisch oder experimentell identifiziert werden. Das gesamte Vorgehen zur Bestimmung des Modells wird dabei als Motorvermessung bezeichnet. Eine Vermessung beinhaltet den Versuchsplanungsprozess (DoE, design of experiments), die eigentliche Messung am Prüfstand sowie den Modellbildungsprozess. Dabei kann zwischen einer stationären und einer dynamischen Vermessung unterschieden werden. Bei der stationären Vermessung wird ein Messplan statisch vermessen, indem jeder einzelne Messplanpunkt angefahren und bis zum Erreichen des stationären Zustandes aller relevanten Größen (z.B. Drücke, Temperaturen, etc.) ausgeharrt wird. Da die eigentliche Messdatenerfassung nur im stationären Zustand erfolgt, bleibt ein Großteil der Messung hierbei ungenutzt. Im Fall einer dynamischen Vermessung erfolgt die Vermessung des Messplans transient mit einem geeigneten Anregungssignal. Die Messdatenerfassung erfolgt dabei durchgehend, ohne dass die Ausgangsgrößen den stationären Zustand erreichen müssen. Basierend auf den sich aus der Messung ergebenden Messdaten können dann verschiedene statische oder dynamische Modelle identifiziert werden.

Versuchsplanung

Die Versuchsplanung erfolgt basierend auf den Anforderungen an das Modell und den sich vom Motor am Prüfstand ergebenden Randbedingungen. Im Falle der stationären Vermessung erfolgt die Versuchsplanung durch die Auswahl geeigneter Kombinationen an Aktoreinstellungen innerhalb des zulässigen Stellbereichs des Motors.

Der zulässige Stellbereich der Aktoren wird dabei als Variationsraum bezeichnet. Die Bestimmung des Variationsraums ist für einen automatisierten Betrieb zwingend erforderlich, um eine Schädigung des Motors am Prüfstand zu vermeiden. Dies kann vorneweg erfolgen, indem mithilfe eines zentral zusammengesetzten Messplans [50] oder eines iterativen Vorgehens [128] die

minimal und maximal zulässigen Aktorkombinationen ermittelt werden. Neuere Ansätze integrieren die Bestimmung des Variationsraums in den eigentlichen Vermessungsprozess [151].

Eine gängige Methode zur Auswahl der zu vermessenden Stellgrößenkombinationen innerhalb des zulässigen Variationsraums erfolgt mithilfe des D-optimalitäts Kriteriums [81]. Mithilfe eines D-optimalen Messplans kann dann die Vermessung am Prüfstand erfolgen [20, 50, 130]. Weitere Möglichkeiten zur Auswahl sind V-optimale sowie raumfüllende Messpläne. Möglichkeiten zur Sortierung der zu vermessenden Stellgrößenkombinationen werden in [138] zusammengefasst.

Der sich ergebende Messplan wird stationär vermessen, indem die Aktoren sprungförmig zu den sich ergebenden Kombinationen verstellt werden bis der stationäre Zustand erreicht ist. Sind die Dynamiken des Prozesses schnell, können ebenso quasistationäre Vermessungsmethoden zum Einsatz kommen [144, 131, 19].

Bei der dynamischen Versuchsplanung muss zusätzlich ein dynamisches Anregungssignal festgelegt werden, welches die ausgewählten Kombinationen ausreichend anregt. Dabei können Signale eingesetzt werden, welche ein kontinuierliches Spektrum an Amplituden aufweisen (Chirp [12]) oder Signale mit nur zwei diskreten Amplituden (Sprung, Rechteck, PRBS [142, 70] und GRBS [174]). Durch die Modulation der sich aus dem Messplan ergebenden Stellgrößenamplituden auf die Sprünge der Signale ergeben sich amplitudenmodulierte Signale, deren Anzahl an Signalsprüngen sich aus den im Messplan vorhandenen Stellgrößenkombinationen ergibt. Zum Einsatz kommen dabei APRBS [70, 138], AGRBS [173] und APRMS [137].

Modellbildung

Bei der Modellbildung von Verbrennungsmotoren kann zwischen kurbelwinkelsynchronen Modellen und Mittelwertmodellen unterschieden werden. Kurbelwinkelsynchrone Modelle bilden dabei die zylinderinternen Vorgänge physikalisch nach, während Mittelwertmodelle die Ausgangsgrößen über ein Arbeitsspiel gemittelt abbilden. Bei Mittelwertmodellen wird zwischen lokalen und globalen sowie stationären und dynamischen Modellen unterschieden [146], wobei aus dynamischen Modellen das stationäre Prozessverhalten bestimmt werden kann [140, 131, 90].

Vermessungsstrategien

Ein Überblick über verschiedene Möglichkeiten zur Vermessung und Identifikation von Verbrennungsmotormodellen geben [69, 175]. Klassische Offline-Ansätze wie z.B. [55], [96], [141] und [146] vermessen den Motor am Prüfstand, um im Anschluss dann eine stationäre Optimierung am Bürorechner vorzunehmen. Neue Methoden kombinieren den Prüfstandsbetrieb mit der Datenauswertung zu Online-Methoden mit dem Ziel der vollautomatisierten Motorkalibrierung direkt am Prüfstand [20, 50]. In [109] ist eine Online-Strategie beschrieben, bei der die stationäre Optimierung der Kennfelder online erfolgt. Dabei werden Modelle für den Kraftstoffverbrauch, HC Emissionen und die Klopfgrenze genutzt. Darauf aufbauend entwickelte BMW

den *mbminimize*-Algorithmus [123, 86, 153, 84], welcher eine online Einzel-Optimierung von Kennfeldern mithilfe von MLP-Netzmodellkomitees durchführt.

Sowohl für Offline- als auch Online-Strategien gibt es heutzutage verschiedene kommerzielle Tools, welche in unterschiedlichem Maße die unterschiedlichen Aspekte der Vermessung vereinen. Diese unterstützen den Applikateur und Prüfstandfahrer, entweder offline am Bürorechner oder online am Prüfstand. Im Einsatz sind dabei die *Model-Based Calibration Toolbox* von MathWorks [133], *ASCMO* von ETAS [94], *CAMEO* von AVL [17], die *Easy DoE ToolSuite* von IAV [11] und die *TOPexpert Suite* von FEV [135].

Neuere Ansätze vereinen den offline und online Einsatz, um dadurch bereits bei der Vermessung am Prüfstand vereinzelt Anpassungen vorzunehmen. Hierzu zählen die *iPROCEDURE ONLINE ADAPTIVE DOE*-Toolbox von AVL [5], welche eine automatische Anpassung des Messplans für einfache Regressionsmodelle und die *KASIO*-Toolbox von Kratzer [14], welche eine offline und online Mehrzieloptimierung von Gauß'schen Prozessmodellen ermöglicht.

1.1.4 Modellbasierte Sollwert-Optimierung

Die Optimierung erfolgt gleichermaßen als stationäre Optimierung von Sollwerten für Steuerungen und Regelungen des Motorsteuergeräts sowie als dynamische Optimierung des Stellgrößenverlaufs durch optimale Einstellung der Steuerungs- und Regelungsparameter. Dabei werden Offline-Ansätze, welche die Optimierung im Anschluss an die Vermessung und Modellbildung durchführen [55, 96, 141] sowie Online-Ansätze, welche eine Optimierung direkt im Betrieb am Prüfstand durchführen [109], genutzt. Die stationäre Optimierung ist seit vielen Jahren Gegenstand der Forschung. Dabei wird zwischen einer lokalen Optimierung, bei der die Motorarbeitspunkte einzeln betrachtet werden, und einer globalen Optimierung unterschieden, bei der mehrere Arbeitspunkte gleichzeitig betrachtet werden [117, 140, 146]. Eine Kombination der lokalen und globalen Optimierung wird in [54] vorgestellt. Dabei wird in einer inneren Schleife eine lokale Optimierung und in einer äußeren Schleife eine dynamische Zyklussimulation zur Überprüfung der Einhaltung von Emissionsgrenzen durchgeführt. Basierend auf den Ergebnissen der äußeren Schleife werden die Gewichte der lokalen Optimierung in der inneren Schleife angepasst, wodurch sich eine globale Optimierung ergibt. Bei einer Optimierung können dabei entweder Verbrauch und Emissionen zueinander gewichtet minimiert werden (Mehrzieloptimierung) oder der Verbrauch wird unter Einhaltung von Emissionsgrenzwerten (Optimierung mit Nebenbedingungen) [140] minimiert.

1.2 Zielsetzung und Aufbau der Arbeit

Das Ziel dieser Arbeit ist die Entwicklung einer Online-Methodik, die es ermöglicht, einen signifikanten Teil des Vermessung- und Modellierungsprozesses automatisiert direkt am Prüfstand durchzuführen. Sie soll ein nichtlineares Prozessverhalten sowohl statisch als auch dynamisch identifizieren können, unter Berücksichtigung von mehreren Ein- und Ausgangsgrößen sowie Arbeitspunkten.

Dabei wird davon ausgegangen, dass die Anwendung der erzeugten lokalen und globalen Modelle, wie z.B. der Regelungsentwurf, weiterhin offline erfolgen. Hierfür müssen Algorithmen zur Versuchsplanung und Modellbildung entwickelt werden, welche ausgehend von Ansätzen aus den Bereichen der statistischen Versuchsplanung und experimentellen Modellbildung (Identifikationsmethoden) für einen Online-Einsatz weiterentwickelt werden.

Zur Umsetzung müssen die einzelnen Schritte des Vermessungs- und Identifikationsprozesses effizienter gestaltet werden. Die einzelnen Schritte sind in Bild 1.5 dargestellt. Im klassischen Vermessungsprozess erfolgt dabei ein Wechsel zwischen Prüfstands- und Büroarbeit. Ziel der Online-Vermessung ist es, den gesamten Prozess an den Prüfstand zu verlagern. Der Prozess muss dabei weitgehendst automatisiert und ohne Benutzereingriffe ablaufen, das heißt ohne manuelle Freigabe von einem Schritt zum nächsten.

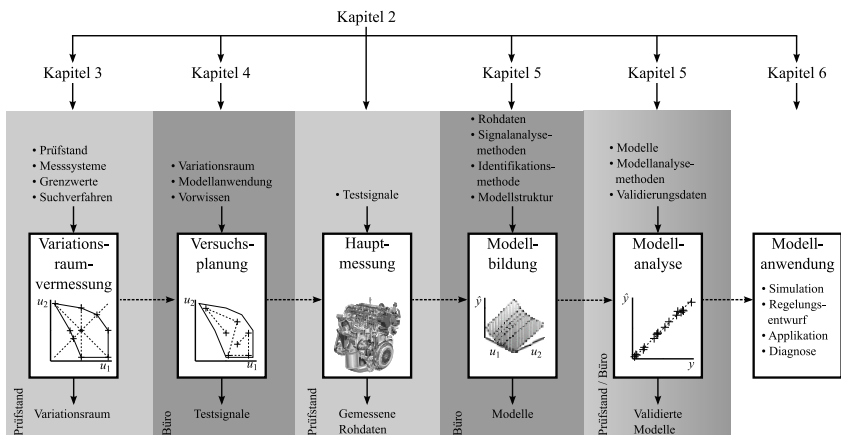


Bild 1.5: Ablauf des Vermessungs- und Modellbildungs-Prozesses mit den Ein- und Ausgängen der jeweiligen einzelnen Schritte.

Ein wichtiger Aspekt ist dabei die Bestimmung des erlaubten Variationsraums der Stellgrößen, um eine Beschädigung des Motors oder ein Betrieb in ungeeigneten Fahrzuständen zu vermeiden. Die Steigerung der Modellgüte bei gleichzeitiger Reduktion der Messzeit kann durch eine gezielte iterative Vermessung im für die Anwendung relevanten Bereich erfolgen. Hierzu werden verschiedene Methoden zur Messplanerstellung entwickelt, die bereits Informationen aus gemessenen Daten und identifizierten Modellen bei der Generierung neuer, zielgerichteter Messbereiche und -punkte nutzt. Die gesamte Methodik soll eine nachträgliche Auswertung mit Unterbrechung des Prüfstandbetriebs unnötig machen. Hierzu muss auch die Validierung bereits in den Prozess integriert werden. Um die Anwendbarkeit der Methodik zu erhöhen und die benötigten Benutzereingriffe zu reduzieren, werden zusätzlich Methoden zur rekursiven Anpassung der Modellstruktur entwickelt. Dadurch kann die erreichbare Modellgüte erhöht werden, und eine Vordefinition der Modellstruktur wird vermieden. Der bisherige Vermessungsprozess soll durch die Verwendung von Online-Methoden verbessert werden, indem sowohl die Ver-

messungszeit an Motorenprüfständen als auch der Aufwand zur Modellierung reduziert werden. Neben dem stationären soll darüber hinaus das dynamische Verhalten des Motors für Laständerungen modelliert werden. Dies ist insbesondere im Hinblick auf kommende sehr geringe Emissionsgrenzen und dynamische Testzyklen von wesentlicher Bedeutung. Die ermittelten Motormodelle sollen für verschiedene Anwendungen anwendbar sein, wie z.B. für optimierte Steuerungen und Regelungen oder modellbasierte Diagnose. Die Praxistauglichkeit und Güte sowie der Zeitaufwand werden dabei im Vergleich zu bisherigen klassischen Methoden bewertet.

Die Hauptbeiträge dieser Arbeit können wie folgt zusammengefasst werden:

- Erstellung eines kombinierten Tools zur Prüfstandautomatisierung sowie Versuchsplanung, welches eine durchgehende, parallele Vermessung und Auswertung ermöglicht.
- Schnelle konkave Variationsraumbestimmung durch Grenzraumregelung.
- Entwicklung neuer Ansätze zur Versuchsplanung mit zielgerichteten Messplänen und dynamischen Kurzzeit-Anregungssignalen.
- Entwicklung von Programmen zur Identifikation von Netzmodellen mit rekursiver Teilungsadaption und richtungsabhängiger Modellierung (ONLOPOMOT).
- Experimentelle Online-Vermessung und Online-Modellbildung von stationären, dynamischen, lokalen sowie globalen Mittelwertmodellen.

Die Struktur dieser Arbeit orientiert sich an den einzelnen Schritten des Vermessungsprozesses aus Bild 1.5.

Die Integration der verschiedenen Online-Verfahren in eine Gesamtmethodik zur Vermessung eines Prüflings am Prüfstand, ausgehend von der Bestimmung des zulässigen Variationsraums bis hin zur Anwendung zur Kennfeldoptimierung, wird in *Kapitel 2* beschrieben. Die hierfür notwendige Anforderungen an den Prüfstand werden dargestellt. Des weiteren wird die Umsetzung der Echtzeitautomatisierung aufgezeigt, bei der die Aktoren in Echtzeit gestellt werden können. Daraus werden verschiedene Verstellstrategien abgeleitet, welche eine schnelle Bestimmung des zulässigen Versuchsraums sowie eine automatische Bestimmung von stationären Messpunkten mithilfe einer an den Prozess angepassten Stationärwerkerkennung ermöglichen.

Der erste Schritt beinhaltet die Bestimmung des zulässigen Stellbereichs aller Aktoren und Darstellung als Variationsraum. Die hierfür eingesetzten Methoden werden in *Kapitel 3* beschrieben. Die eigens entwickelte Methode zur schnellen Ermittlung und dennoch einfachen Beschreibung von zulässigen Variationsräumen wird vorgestellt. Mithilfe derer können gleichermaßen lokale wie globale konkave Variationsräume beschrieben werden.

Die bei der Vermessung darauf folgende Versuchsplanung wird in *Kapitel 4* dargestellt. Die in der Online-Methodik umgesetzte Online-Versuchsplanung generiert innerhalb des zulässigen Variationsraums, basierend auf verschiedenen Ansätzen, zielgerichtet neue Messpunkte. Dabei

werden sowohl die tatsächlich gemessenen Messgrößenvarianzen als auch verschiedene Modelleigenschaften sowie Modelloptima genutzt. Ausgehend von den Messpunkten werden dann dynamische Anregungssignale entwickelt, welche für eine iterative Kurzzeit-Anregung geeignet sind.

Im Anschluss an die eigentliche Messung am Prüfstand erfolgt die Modellbildung und die Modellanalyse. In *Kapitel 5* werden notwendige Erweiterungen der klassischen Identifikation, welche sich durch den Einsatz im Rahmen der Online-Methodik ergeben, beschrieben. Der wesentliche Unterschied zwischen Offline und Online-Verfahren im Hinblick auf Modellanalyse und das Bias-Varianz-Dilemma wird diskutiert sowie verschiedene Möglichkeiten zur Erhöhung der Modellgüte werden vorgestellt. Neben der Möglichkeit der automatischen iterativen Adaption der Regressoren und Teilung bei Netzmodellen kann auch eine richtungsabhängige Modellierung zum Einsatz kommen.

Die Online-Vermessung wird abschließend anhand eines Anwendungsbeispiels in *Kapitel 6* verdeutlicht, bei der die lokale Vermessung mit dem Ziel der Parametrierung von Sollwert-Kennfeldern automatisiert erfolgt.

Die vorgestellten Methoden und Ergebnisse werden in *Kapitel 7* zusammengefasst und ein Ausblick auf weitere Forschungsarbeit wird gegeben.

2 Gesamtstrategie zur Online-Vermessung

Die in dieser Arbeit entwickelten Methoden sowie deren Einbettung in eine durchgehende Gesamtstrategie stellen verschiedene neue Anforderungen an den zugrundeliegenden Prüfstand sowie dessen Automatisierung, welche im Folgenden vorgestellt werden. Wesentlicher Aspekt ist dabei die Kombination der Variationsraumbestimmung, der Versuchsplanung und der Identifikation in eine Gesamtstrategie zur Online-Vermessung. Alle im Rahmen dieser Arbeit entwickelten Algorithmen wurden in die Toolbox *OTOM* (online target oriented measurement) und *ORTEM* (online real-time engine measurement) integriert. Erstere basiert auf MATLAB und entspricht der Weiterentwicklung von Offline-DoE-Tools während letztere eine auf Simulink und ControlDesk basierende Prüfstandsautomatisierung ist. Durch die Kombination beider wird eine automatisierte Online-Vermessung an einem Prüfstand mit Echtzeithardware umgesetzt.

Bislang erfolgt die Motorvermessung in der Praxis meist in einzelnen Schritten sequentiell und mit einem zeitaufwändigen örtlichen Wechsel von Vorbereitung am Arbeitsplatz sowie Durchführung der Messungen am Motorenprüfstand. Bei dem sequentiellen Vorgehen ergeben sich Rückschritte, in denen ein Applikateur zur Verbesserung der Modellbildung die Vermessung adaptiert und diese erneut durchgeführt wird. Diese Rückschritte haben zur Folge, dass sie zum gezielten Vermessen das nötige Wissen eines Applikateurs und die Verwendung verschiedener Softwaretools benötigen sowie aufgrund des wiederholten Wechsels zwischen Prüfstands- und Arbeitsplatzrechner zu erhöhten Stillstandzeiten und erhöhten Messzeiten am Prüfstand führt.

Eine deutliche Verbesserung mit Hinblick auf den Vermessungsaufwand, den Auswertungsaufwand und die spätere Anwendung lässt sich erreichen, wenn während der einzelnen Schritte der Vermessung und Modellbildung der Motor und Rechner ständig stell- und messseitig gekoppelt sind. Dieses wird als *Online*-Betrieb bezeichnet. Bei einer Online-Motorvermessung muss der Applikateur/Prüfstandsfahrer nicht mehr manuell den Übergang von einem Schritt zum nächsten durchführen, sondern der Prüfstandsrechner nutzt automatisch das Ergebnis des vorherigen Schrittes um selbstständig weiter im Ablauf vorzugehen. Hierfür müssen alle relevanten Vorgaben (z.B. Vorwissen, anzuwendende Identifikations- / Analysemethoden) bereits vor Beginn der eigentlichen Vermessung gemacht werden.

Durch die Kopplung der Prüfstandsautomatisierung mit dem Auswertungs-Tool können verschiedene Vorteile und Verbesserungen erreicht werden, die den gesamten Modellbildungsprozess effektiver machen und zu einer Reduktion der Prüfstandszeit führen. Diese lauten wie folgt:

- Iterative Vorgehensweise:
 - Steigerung der Modellgüte durch gezielte Vermessung in relevanten Bereichen für die Anwendung.

- Reduktion der Messzeit für gewünschte Modellgüte.
- Anwendungsspezifische Vermessung.
- Reduzierte Benutzereingriffe während eines Ablaufs:
 - Schnelle, automatisierte Betriebsbereichbestimmung und dessen Berücksichtigung im weiteren Vermessungsverlauf.
 - Automatisierte Aufeinanderfolge von einem Schritt zum nächsten Schritt (z.B. von Hauptmessung zu Modellbildung) – geringe Stillstandszeiten des Prüfstands.
 - Motor bleibt durchgehend im gleichen konditionierten Betriebszustand – keine wiederholten Warmlaufzeiten nötig.
- Direkte Datenaufbereitung, Modellbildung und Validierung am Prüfstand. Eine nachträgliche Überprüfung der Modellgüte (im Büro) entfällt.
- Entfall des Wechsels von Rechnern und Tools zwischen Büro und Prüfstand.
- Direkter Übergang zum Entwurf von Steuerungen und Regelungen.

2.1 Prüfstandsanforderungen

In Bild 2.1 ist ein exemplarischer Aufbau eines Prüfstands dargestellt, welcher für Online-Methoden geeignet ist.

Der Prüfstand benötigt eine entsprechende Prüfstandsautomatisierung und einen Prüfstandsrechner zur Eingabe der Stellgrößen und Speicherung der gemessenen Daten. Der hohe Automatisierungsgrad bei einer Online-Vermessung erfordert eine sichere Kommunikation aller beteiligten Systeme und eine Überwachung der Funktionen und kritischer Grenzen. Dies kann mithilfe eines Rechnersystems mit entsprechender Echtzeithardware erreicht werden. Hierbei kooperieren alle am Prüfstand zum Einsatz kommenden Systeme über einen CAN-Bus miteinander. Neben den klassischen Rechnern zur Prüfstandssteuerung und Kalibrierung bilden die Rechner zur Online-Vermessung und Online-Auswertung die zusätzlichen Bestandteile, die bei der Online-Methodik zum Einsatz kommen.

Für die stationäre und vor allem dynamische Online-Modellbildung wird ein Prüfstand benötigt, der alle Aktoren dynamisch ohne wesentlichen Zeitverzug stellen kann, sowie über schnelle dynamische Sensoren mit kleinen Zeitkonstanten und kleinen Totzeiten verfügt. Für die Berücksichtigung von Verbrennungsgrößen, wie z.B. Schwerpunktlage oder Zylinderspitzen drücke, muss zusätzlich ein Echtzeit-Druckindiziersystem existieren. Weiterhin muss eine schnelle Datenaufzeichnung gewährleistet sein.

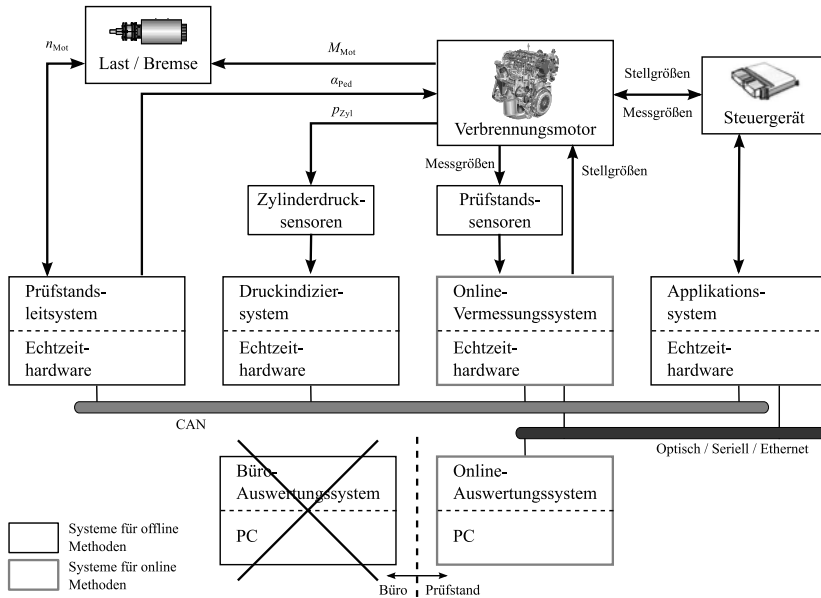


Bild 2.1: Prüfstandsschema und Signalflüsse für den Einsatz von Online-Methoden an Motorenprüfständen.

2.2 Iterativer Ablauf der Gesamtstrategie

Durch die Kombination der automatischen Online-Vermessung und der Online-Auswertung wird eine Parallelisierung des iterativen Vermessungsablaufs erreicht (siehe Bild 2.2). Dabei wird während der Datenauswertung und Modellbildung/-analyse auf dem Prüfstands-PC bereits parallel der nächste Messplan selbsttätig von der Echtzeithardware am Prüfstand vermessen. Die Zeit, welche für eine nachträgliche Auswertung der Messdaten im Büro benötigt wird, kann dadurch eingespart werden. Wesentlicher Vorteil des sich ergebenden iterativen Batch-Vorgehens ist, dass viele der in den letzten Jahren entwickelten neuen Methoden der Identifikation eingesetzt werden können.

Der sich ergebende Ablauf ist Folgender:

1. Iteration: Es wird ein 1.Messplan zur Bestimmung des Variationsraums erstellt (z.B. Hüllraster, siehe Abschnitt 3.3) und vermessen.
2. Iteration: Nach Abschluss der 1.Iteration werden die Messdaten konvertiert. Aus den Messdaten der Variationsraumvermessung wird die mathematische Beschreibung des Variationsraums bestimmt. Basierend auf dem Variationsraum wird ein modellunabhängiger Online-Messplan Nr.2 erstellt und die Vermessung dessen wird gestartet. Im Falle der dynamischen Vermessung wird ein zusätzlicher modellunabhängiger Online-Messplan

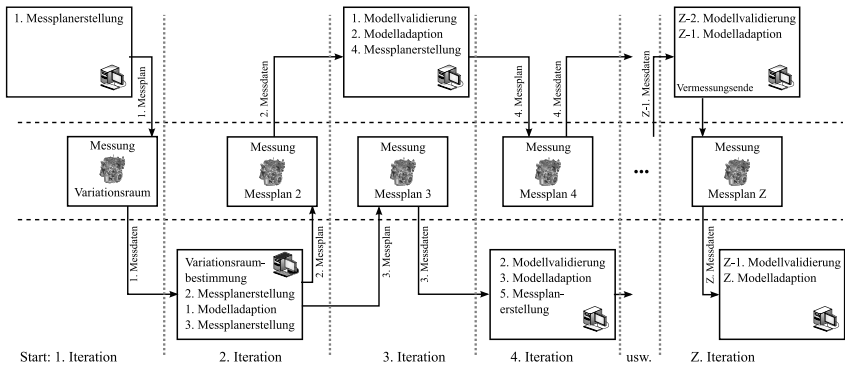


Bild 2.2: Iterativer Ablauf der Online-Vermessung von Verbrennungsmotoren mit wechselnder Vermessung und Auswertung sowie integrierter Validierung.

unter Berücksichtigung des Variationsraums als 2. Messplan erstellt. Im Gegensatz dazu erfolgt bei der stationären Vermessung eine Identifikation und Analyse des 1. Modells mit den Messdaten aus Iteration 1 sowie die Erstellung eines modellabhängigen Online-Messplans unter Berücksichtigung des Variationsraums als 2. Messplan.

3. Iteration: Nach Abschluss der 2. Iteration wird die Vermessung des 3. Messplans gestartet. Parallel zur laufenden Vermessung erfolgt die Konvertierung der Messdaten sowie Identifikation und Analyse des 2. Modells anhand aller bisher vorhandenen Messdaten. Im Fall der stationären Vermessung wird zusätzlich das 1. Modell mit dem Messdaten der 2. Iteration validiert. Abschließend wird ein modellabhängiger Online-Messplan als 4. Messplan unter Berücksichtigung des Variationsraums erstellt.
4. Iteration: Nach Abschluss der 3. Iteration wird die Vermessung des 4. Messplans gestartet. Parallel zur laufenden Vermessung erfolgt die Konvertierung der Messdaten, die Validierung des 2. Modells mit den Messdaten der 3. Iteration, sowie die Identifikation und Analyse des 3. Modells mit allen bisher vorhandenen Messdaten. Abschließend wird ein modellabhängiger Online-Messplan als 5. Messplan unter Berücksichtigung des Variationsraums erstellt.
- ⋮
- Z. Iteration: Nach Abschluss der $Z - 1$. Iteration wird die Vermessung des Z . Messplans gestartet. Parallel zur laufenden Vermessung erfolgt die Konvertierung der Messdaten, die Validierung des $Z - 2$. Modells mit den Messdaten der $Z - 1$. Iteration sowie die Identifikation und Analyse des $Z - 1$. Modells mit allen bisher vorhandenen Messdaten. Aufgrund des erreichten Abbruchkriteriums (min. Modellgüte oder max. Iterationen) wird kein neuer Messplan generiert, sondern das Ende der aktuellen Iteration abgewartet. Nach Abschluss der Z . Iteration wird die iterative Online-Vermessung beendet, indem

die Konvertierung der Messdaten sowie die Identifikation und Analyse des Z. Modells mit allen vorhandenen Messdaten stattfindet.

Bei der Parallelisierung werden somit während einer laufenden Messung gleichzeitig die Daten der vorhergehenden Messung konvertiert und aufbereitet sowie das Modell gebildet und analysiert. Speziell im Fall der dynamischen Modellbildung, bei der kontinuierlich Daten aufgezeichnet werden, muss die Auswertung schneller abgeschlossen sein als die Messung, um ein zu langes Warten in einem stationären Punkt zu vermeiden. Durch die Parallelisierung können zusätzlich in die Gesamtmethodik Validierungsmessungen integriert werden, um die Genauigkeit der bestimmten Modelle zu überprüfen. Mit den neuesten Daten wird deshalb zuerst das Modell der vorhergehenden Messung validiert, um die Güte beurteilen zu können. Erst im Anschluss daran wird mit allen bisherigen Daten das neue Modell gebildet. Dieses Vorgehen erübrigt die abschließende Validierung mit einem separaten Datensatz, womit weitere Prüfstandszeit eingespart werden kann.

2.3 Schnittstellen und Struktur der Gesamtstrategie

Der schematische Zusammenhang zwischen den einzelnen Teilsystemen ist in Bild 2.3 dargestellt. Die gesamte Online-Methodik besteht aus der Online-Vermessung auf der Echtzeithardware kombiniert mit der Online-Auswertung auf einem PC.

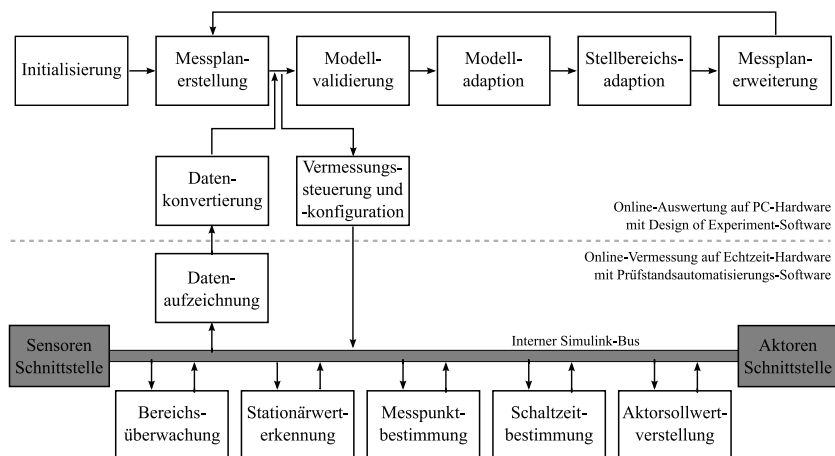


Bild 2.3: Struktur der Interaktion zwischen der Echtzeit-Vermessung und der Online-Auswertung.

Die Online-Auswertung nutzt die von der Online-Vermessung zur Verfügung gestellten Daten von Anfang an, um im laufenden Betrieb eine Datenkonvertierung, Datenanalyse, experimen-

telles Modellbildung, Modellanalyse und Modellvalidierung durchzuführen. Die Datenverarbeitung erfolgt also im Batch-Modus, wobei die Schritte der Online-Auswertung überwiegend rekursiv erfolgen. Nach erfolgter Modellvalidierung übernimmt die Online-Auswertung zusätzlich die Parametrierung der nächsten Vermessung, indem ein neuer Messplan erstellt, an das Echtzeitsystem übergeben und die folgende Vermessung gestartet wird. Somit übernimmt das Auswertungssystem auch die zentrale Steuerung des gesamten Prüfstandes während der laufenden Online-Vermessung.

In einer dSPACE Echtzeitumgebung bildet MLIB, einer Matlab Bibliothek zur Steuerung von dSPACE-Systemen, die Schnittstelle zwischen der Auswertung und Echtzeit-Vermessung. Mithilfe von Funktionen, welche von MLIB zur Verfügung gestellt werden, können die für die Echtzeit-Vermessung benötigten Variablen an die Echtzeithardware übergeben und die Erfassung von Messdaten gesteuert werden, ohne den Betrieb des Echtzeitsystems zu unterbrechen. Ferner werden die Daten mit DSConvert konvertiert, einem dSPACE eigenen Tool, welches die vom Echtzeitsystem aufgezeichneten Daten in entsprechende Matlab-Strukturen umwandelt.

Die Messdatenerfassung erfolgt hierbei je nach Modellbildungs- bzw. Vermessungsart. Bei einer dynamischen Vermessung werden die Messdaten kontinuierlich erfasst, wobei ein direktes Downsampling durch Anpassung der Triggerung der Datenaufzeichnung geschieht. Wird stationär vermessen, wird die Datenerfassung im stationären Zustand aufgezeichnet. Hierzu wird ein Intervall vorgegeben, über welches die abgetasteten Größen gemittelt werden. Nach einer Mittelung werden die Mittelwerte der aufzuzeichnenden Größen gespeichert und der nächste Messplanpunkt angesteuert.

Die Online-Vermessung mit einer Echtzeithardware übernimmt die Kommunikation mit der Prüfstandsumgebung über die entsprechenden A/D- und I/O-Karten. Sie gliedert sich in folgende Teilsysteme:

- Sensoren-/Aktoren-Schnittstelle: Echtzeit-Schnittstelle zu allen benötigten Aktoren und Sensoren
- Bereichsüberwachung: Beinhaltet die Bereichsüberwachung zur Erkennung von harten und weichen Grenzen
- Stationärwerterkennung: Beinhaltet die Stationärwerterkennung aller relevanten Messgrößen
- Verstellgeschwindigkeitsadaption: Beinhaltet die Adaption der Verstellgeschwindigkeit du/dt in Abhängigkeit des Grenzabstandes bei der Variationsraumvermessung
- Messplanpunkt- und Schaltzeitbestimmung: Beinhaltet die Bestimmung der aktuell anzusteuernden Stellgrößenkombination sowie den Schaltzeitpunkt, ab welchem der nächste Messplanpunkt angesteuert werden soll
- Aktorsollwertgenerierung: Beinhaltet die Echtzeitgenerierung der Aktorsollwerte

- **Datenaufzeichnung:** Beinhaltet die angepasste Aufzeichnung der Messgrößen auf die Festplatte

Durch die Kombination sowohl der DoE- als auch der Prüfstandsautomatisierungs-Software in eine Gesamtmethodik werden verschiedene Verbesserungen gegenüber dem herkömmlichen Vorgehen ermöglicht. Durch diese Kombination am Motorenprüfstand erfolgt eine Integration, wodurch der gesamte Prozess der Motorvermessung und Modellbildung automatisch abläuft. Der Ablauf sowie die Steuerung der Aktoren, Datenaufzeichnung und Auswertung erfolgen dabei automatisiert, im Idealfall ohne Benutzereingriffe im laufenden Betrieb.

2.3.1 Online-Vermessungssystem *ORTEM*

Ein herkömmlicher PC ist nur eingeschränkt echtzeitfähig. Ein vom PC an die Aktoren vorgegebener Messplan kann wegen der benötigten Übertragungslaufzeit von PC zum Echtzeitsystem nicht schnell bzw. genau genug abgefahren werden. Um eine in Echtzeit ablaufende Online-Vermessung zu realisieren, muss diese als kompilierter Code direkt auf einer entsprechenden Echtzeithardware (von z.B. dSPACE, ETAS, NI etc.) ausgeführt werden.

Die Online-Vermessung *ORTEM* (online real-time engine measurement) ist in Bild 2.4 vereinfacht dargestellt. Die sich aus dem Messplan ergebenden Stellgrößenkombinationen werden als einzelne Vektoren für jeden Aktor an das Echtzeit-System übergeben und über einen internen Zähler, der bei Erreichen des Schaltzeitpunktes hochzählt, angesteuert.

Der aktuell angefahrne Messplanpunkt definiert sich über den gewünschten Aktorsollwert $u_{\text{Sollwerte}}$, die sich aus der gewünschten Stellzeit (Dauer bis zum Erreichen des Messplanpunktes) ergebende Aktorverstellgeschwindigkeit $u_{\text{Steigungen}}$ sowie den Schaltzeitpunkten $u_{\text{Schaltzeitpunkte}}$. Die Aktorverstellgeschwindigkeit u_{Steigung} ergibt sich aus den Eigenschaften der Steller bzw. der Steigung der definierten Anregung. Der Schaltzeitpunkt $u_{\text{Schaltzeitpunkt}}$ bestimmt sich für jeden Messplanpunkt automatisch aus der aktuellen fortlaufenden Systemzeit t_{Echtzeit} und der Summe aus Stellzeit, Einschwingzeit und Mittelwertszeit bei der stationären Vermessung bzw. Stellzeit und Taktzeit bei der dynamischen Vermessung (siehe Bild 4.2).

Die adaptive Sollwertverstellung der Online-Vermessung im Echtzeitsystem basiert auf Sprüngen, welche mit einer variablen Änderungsgeschwindigkeit du/dt versehen sind. Dadurch kann neben der sprungförmigen ($du/dt = \infty$) auch die rampenförmige ($0 < du/dt < \infty$) Aktorverstellung umgesetzt werden.

Die Vermessung erfolgt im Gegensatz zu anderen Systemen in Echtzeit, indem alle notwendigen Größen des Messplans ($u_{\text{Sollwerte}}$, $u_{\text{Steigungen}}$, $u_{\text{Schaltzeitpunkte}}$) zu Beginn der Iteration als Vektoren mit der Länge $i = N$ vorgegeben werden. Das Echtzeitsystem kann dadurch selbständig bei Erreichung des Schaltzeitpunktes das nächste Element im Vektor (entspricht dem nächsten Messplanpunkt) als Sollwert an den Aktoren ausgegeben, ohne das eine Übertragungslaufzeit von PC zum Echtzeitsystem benötigt wird zur Ausführung des Stellbefehls.

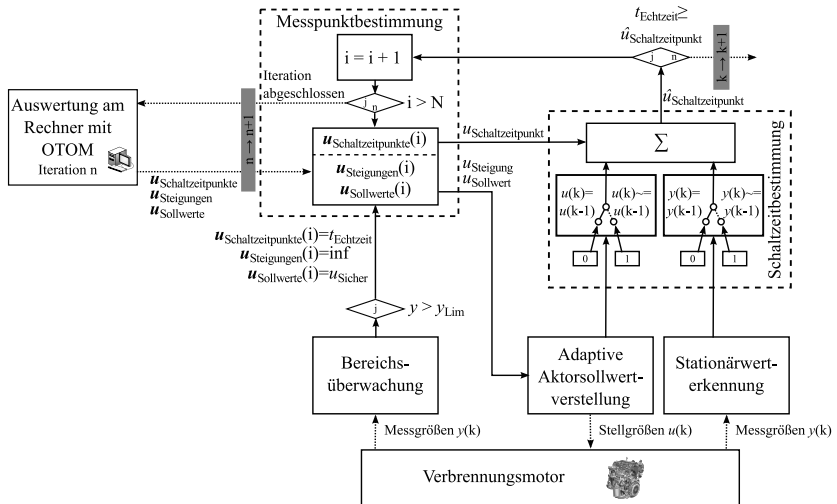


Bild 2.4: Schematischer Ablauf der Echtzeit-Online-Vermessung ORTEM. Die Laufvariablen k und i entsprechen hierbei den diskreten Zeitschritten bzw. den Messplanpunkten. Der Schaltzeitpunkt $u_{\text{Schaltzeitpunkt}}$, bei dem der nächste Messplanpunkt $i = i + 1$ angesteuert wird, wird solange um ein Abtastschritt nach hinten verschoben, bis die Sollwertverstellung abgeschlossen ist und die Stationärwerverkennung alle Größen als eingeschungen detektiert.

Die Online-Vermessung beinhaltet neben der eigentlichen Verstellung aller Aktoren zusätzlich eine Online-Sensorüberwachung, welche sich aus einer Bereichsüberwachung und einer Stationärwerverkennung zusammensetzt.

Die Bereichsüberwachung nutzt Messgrößenlimits, um Grenzverletzungen zu detektieren. Im Rahmen der Online-Methoden wird eine Definition von harten und weichen Grenzen genutzt, welche unabhängig von der entsprechenden Messgröße ist. Hierbei sind weiche Grenzen von Messgrößen Messwerte, welche prinzipiell eingehalten werden sollten, wobei angenommen wird, dass ein geringes kurzfristiges Überschreiten (z.B. Überschwingen der Messgröße) noch keine Schädigung zur Folge hat und somit zugelassen wird. Wird bei der stationären Vermessung die vorgegebene weiche Grenze einer Ausgangsgröße erreicht, wird die Verstellung automatisch angehalten, der stationäre Zustand aller für die Vermessung relevanten Größen abgewartet, die aktuelle Aktorposition gespeichert und der nächste Messplanpunkt wird angesteuert. Hierdurch wird der Variationsraum im laufenden Betrieb kontinuierlich erfasst, sodass er iterativ angepasst werden kann. Eine harte Grenze hingegen stellt ein nicht zu überschreitendes Maximum bzw. Minimum der Messgröße dar. Wenn eine harte Grenzverletzung durch die Bereichsüberwachung erkannt wird, erfolgt, nach einer Abspeicherung der Aktorpositionen, ein sofortiger Rückzug aller Aktoren am Prüfstand in einen definierten, sicheren Betriebspunkt. Nach Erreichen des stationären Zustandes in dem sicheren Betriebspunkt wird der nächste

Messplanpunkt angefahren. Die Reaktion auf die harte Grenzverletzung geschieht dabei automatisch, ohne dass ein Eingriff des Benutzers nötig ist.

Während die Bereichsüberwachung immer im Hintergrund aktiv ist, kommt die Stationärwerterkennung bei der Bestimmung des Variationsraumes und bei der Vermessung von stationären Modellen zum Einsatz. Im Gegensatz zum klassischen Vorgehen bei der stationären Modellbildung, bei dem eine pauschale, gleichbleibende Einschwingzeit angenommen und abgewartet wird, bevor die stationäre Datenerfassung gestartet wird, erfolgt bei der Online-Methode die Bestimmung einer messgrößen- und messpunktindividuellen Einschwingzeit mit der Echtzeithardware. Dies ist besonders dann interessant, wenn die Einschwingzeiten der Ausgangsgrößen verschieden sind. Sobald der eingeschwungene Zustand der relevanten Messgrößen erkannt wird, erfolgt eine automatische Datenaufzeichnung mit einer Mittelung über die definierte Mittelwertszeit sowie ein automatischer Übergang zum nächsten Messplanpunkt.

Die Stationärwerterkennung kann durch den Vergleich des Messwerts zu verschiedenen Zeitpunkten realisiert werden. Die Differenz zwischen den verglichenen Messwerten darf nicht größer als die doppelte Standardabweichung 2σ der entsprechenden Messgröße sein, damit der stationäre Zustand erkannt wird. Durch die Kombination eines kurzfristigen Vergleichs (Intervall L_k) und eines mittelfristigen Vergleichs (Intervall L_m) wird auch die Erkennung von ein- und überschwingenden Verläufen erreicht.

Die Parametrierung der Erkennungsgrenze 2σ sowie der Intervalle L_k und L_m erfolgt in Abhängigkeit von der Summenzeitkonstante T_Σ (entspricht der Summe aus Sensorzeitkonstante und Prozesszeitkonstante) und der Standardabweichung des gemessenen Signals. Dabei können die Sensorzeitkonstanten ca. Millisekunden (NO_x -Sensor) bis Sekunden (MicroSoot-Sensor) und die Prozesszeitkonstanten wenige Sekunden (NO_x) bis mehrere Minuten (Temperaturen) betragen. Bei Annahme eines PT_1 -Verhaltens kann die Summenzeitkonstante näherungsweise als T_1 ($= 63\%$ des Endwerts) angenommen werden. Für die Intervalle ergibt sich:

$$L_m = T_1; \quad L_k = \frac{L_m}{3}. \quad (2.1)$$

Mithilfe der so erfolgenden Parametrierung ergibt sich eine ausgangsgrößenindividuelle Einschwingzeit, welche die Vermessungszeit wesentlich reduziert, da keine pauschale Einschwingzeit bis zum Erreichen des eingeschwungenen Zustands aller Messgrößen mehr abgewartet werden muss, z.B. 3 min. Im Anschluss an die Erkennung des eingeschwungenen Zustandes wird die Datenaufzeichnung automatisch gestartet, und die Messwerte werden für eine vorzugebende Mittelungszeit aufgezeichnet.

In Bild 2.5 sind für den einschwingenden Beispielprozess aus (D.1) sowohl die Intervalle als auch die Signalverläufe für zwei verschiedenen Sollwertsprünge dargestellt. Es wird deutlich, dass die Erkennung des stationären Zustandes bei dem kleineren Sollwertsprung früher erfolgt, wodurch die Vermessung automatisiert schneller zum nächsten Sollwert übergehen kann.

Wie in Bild 2.6 a) dargestellt, dauert die Vermessung der 10 Messplanpunkte bei einer Verstellung mit Stationärwerterkennung von NO_x und T_3 insgesamt 107 s. Die längste Übergangsfunktion beträgt dabei 20 s. Unter Annahme einer konstanten Schaltzeit müsste diese dadurch

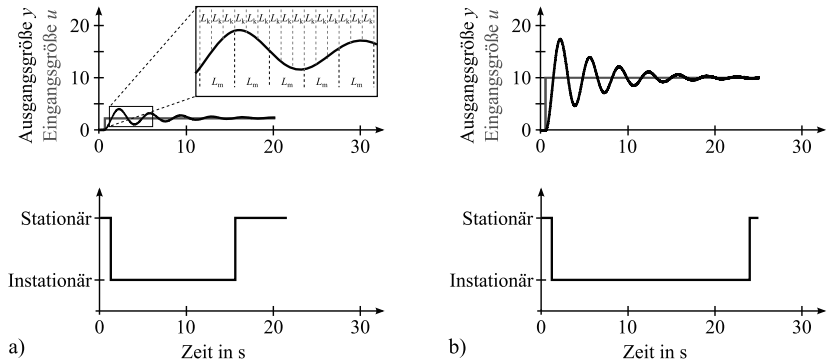


Bild 2.5: Darstellung der Vergleichsintervalle sowie Verlauf der Stationärwerterkennung bei zwei verschiedenen Sollwertsprüngen eines dynamischen Beispielprozesses.

auf mindestens 20 s festgelegt werden. Der daraus resultierende zeitliche Verlauf einer Vermessung ohne Stationärwerterkennung ist in b) dargestellt. Die Vermessung dauert 200 s. Durch den Einsatz der automatischen Schaltzeitadaption bei Erkennung von stationär eingeschwungenen Größen kann die Dauer der Vermessung in diesem Beispiel um ca. 47% reduziert werden.

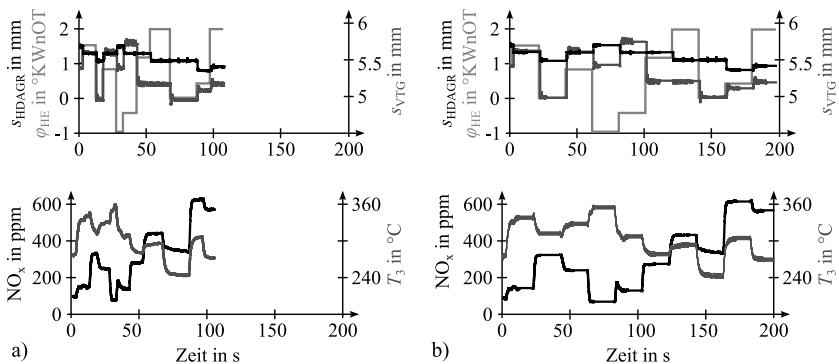


Bild 2.6: Vermessung der Ausgangsgrößen NO_x und T_3 bei Verstellung von s_{HDAGR} , s_{VTG} und ϕ_{HE} mit (a) und ohne (b) automatischer Schaltzeitadaption bei Stationärwerterkennung.

2.3.2 Online-Auswertungssystem OTOM

Das Online-Auswertungssystem *OTOM* (online target oriented measurement) nutzt Messdaten, identifizierte Modelle und Ergebnisse einer Sollwertkennfeldoptimierung zur aktiv lernenden Adaption von Messplänen und Modellen. Mithilfe der neusten Messdaten wird dabei zuerst das

neueste Modell mit adaptierter Struktur identifiziert. Mittels der adaptierten Modelle wird eine modellbasierte Adaption des Messplans durchgeführt. Die Adaption erfolgt iterativ innerhalb des gesamten erlaubten Variationsraumes. Die im Folgenden vorgestellte Gesamtstrategie erreicht eine Verbesserung des Modells in dem später benötigten Bereich in kurzer Zeit, wodurch die erforderliche Prüfstandmesszeit weiter reduziert wird. Sie ist dabei sowohl für lineare und nichtlineare als auch für statische und dynamische sowie lokale und globale Modelle anwendbar.

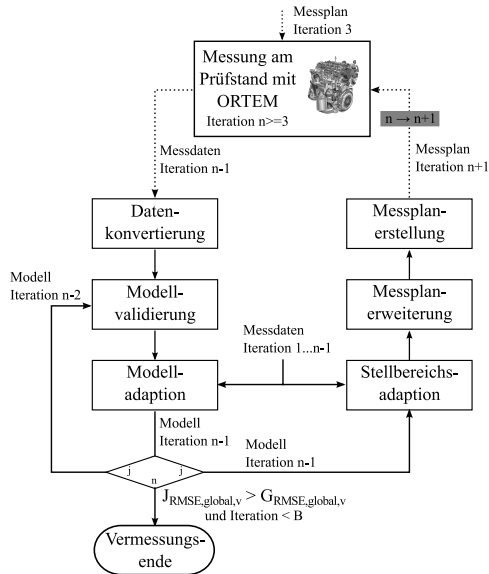


Bild 2.7: Gesamtstrategie zur Adaption des Modells und des Messplans innerhalb des Online-Auswertungssystem *OTOM* ab Iteration ≥ 3 .

In Bild 2.7 ist die sich ergebende Anpassungsstrategie ab der 3. Iteration dargestellt. Sobald neue Messdaten vermessen worden sind, beginnt die Modellanpassung. Im ersten Schritt wird eine Validierung des alten Modells mit den neuesten Messdaten sowie Modelladaption entsprechend 5.4.5 durchgeführt. Falls die Zielgüte sowie die maximale Anzahl an Iterationen Z noch nicht erreicht ist, werden der aktuelle Aktorstellbereich angepasst sowie der Messplan um neu zu vermessende Punkte, wie in Abschnitt 4.3.5 beschrieben, erweitert. Sobald neue Messdaten aufgezeichnet worden sind, beginnt der Ablauf der Modell- und Messplananpassung von neuem. Nicht dargestellt sind die Iterationen 1 und 2, da diese vorneweg mit definierter Messplanart (Hüllraster und DTS) ohne Adaption von Modell und Messplan ablaufen.

Mithilfe der beschriebenen Methodik wird das Modell bei Bedarf iterativ an den Prozess angepasst, indem die Struktur kontinuierlich adaptiert wird. Dadurch kann eine maximale Verbesserung des Modells erreicht werden. Durch die Integration der Validierung in die iterative

Adaption kann außerdem auf eine Validierung im Nachgang mit separaten Messdaten verzichtet werden, da bereits in jeder Iteration eine Validierung und damit Überprüfung des letzten Modells mit neuen Messdaten durchgeführt wird.

2.4 Zusammenfassung

Durch die stell- und messseitige Kopplung der Vermessung mit der Modellbildung kann die Effektivität des Modellbildungsprozesses erhöht werden. Hierfür wird ein Prüfstand benötigt, welcher mit dynamischen Sensoren und Aktoren sowie einem Echtzeit-Rechnersystem ausgestattet ist, um eine schnelle Verstellung sowie Datenaufzeichnung zu ermöglichen.

Das entwickelte Online-Vermessungssystem ORTEM ermöglicht eine echtzeitfähige Verstellung der Aktoren bei gleichzeitiger Limit-Überwachung. Durch die Erweiterung um die automatische Adaption der Aktor-Verstellgeschwindigkeit sowie den automatischen Übergang zum nächsten Messplanpunkt bei Stationärwerterkennung kann die Messzeit reduziert werden. Die Stärke der Reduktion ist dabei abhängig vom Anwendungsfall (den vermessenen Größen sowie der Anzahl an Messpunkten) und beträgt in dem gezeigten Beispiel 50 %.

Das Online-Auswertungssystem OTOM nutzt Messdaten und identifizierte Modelle zur aktiv lernenden Adaption von Messplänen und Modellen. Dabei werden diese solange iterativ angepasst und neue Messdaten vermessen, bis die Zielgüte erreicht ist. Infolgedessen kann sowohl eine maximale Verbesserung des Modells im Hinblick auf Bias und Varianz erreicht werden, als auch eine unnötige Vermessung von Messpunkten ohne wesentliche Prozessinformation vermieden werden, was wesentliche Messzeit einspart.

Die Kombination beider Systeme ermöglicht eine Parallelisierung des iterativen Vermessungsablaufs. Dabei wird während der Datenauswertung und Modellbildung/-analyse auf dem Prüfstands-PC mit OTOM bereits parallel der nächste Messplan selbsttätig von der Echtzeithardware mit ORTEM am Prüfstand vermessen. Während eine Vermessung läuft, kann bereits eine Auswertung durchgeführt werden. Neben dem dadurch entfallenden Wechsel zwischen Arbeitsplatz und Prüfstand ist der gesamte Aufwand, der beim klassischen Vorgehen für das nachträgliche Auswerten im Büro anfällt, nicht nötig.

3 Methoden zur Bestimmung zulässiger Aktor-Stellbereiche

Grundlage einer automatisierten Vermessung stellt die vorgelagerte Bestimmung der zulässigen Aktor-Stellbereiche dar, um sowohl einen sicheren Betrieb des Prüfstandes als auch die Vermessung innerhalb vordefinierter Grenzen sicherzustellen. Um die Bestimmung der zulässigen Stellbereiche im Vergleich zu bisherigen Methoden zu beschleunigen, wird der Grenzraum mithilfe eines speziellen Messplans sowie einer geregelten Aktor-Verstellgeschwindigkeit vermessen. Die so bestimmten Hüllpunkte werden im Anschluss mithilfe einer einfachen, auf konvexen Hüllen basierenden, konkaven Hülle beschrieben. Die neu entwickelte Art zur Beschreibung der Hülle ermöglicht die Interpolation von Hüllen in Zwischenarbeitspunkten sowie den Verzicht auf einen aufwändigen nichtlinearen Modellierungsprozess von Hüllen.

Bei der Verwendung von Online-Methoden läuft der Vermessungsprozess weitgehend automatisiert ab. Nicht alle theoretisch möglichen Stellgrößenkombinationen des Motors resultieren in einem beschädigungsfreien Betrieb. In bestimmten Betriebsbereichen können sich zu hohe Zylinderspitzendrücke oder Temperaturen ergeben, welche zu einer Beschädigung des Motors führen können. Bevor automatisierte stationäre und dynamische Messungen am Prüfstand durchgeführt werden können, müssen die stellbaren Bereiche aller Aktoren bestimmt werden, um den automatisierten Betrieb in nicht zulässigen Fahrzuständen, bzw. im schlimmsten Fall die Beschädigung des Motors, zu vermeiden. Ein manuelles Eingreifen zur Anpassung der Stellgrößen an den Variationsraum ist in einer solchen Methode nicht möglich. Die Bestimmung und Berücksichtigung des erlaubten Variationsraums der Stellgrößen stellt deshalb einen wichtigen und grundlegenden Aspekt der Online-Motorvermessung dar. Zur Bestimmung des Variationsraums wird ein Suchverfahren angewendet, welches ausgehend von einem sicheren Startpunkt die Aktoren solange verstellt bis das Stellbereichsmaximum oder vorgegebene Grenzen der Ausgangsgrößen erreicht werden. Das Suchverfahren ist abhängig von den vorhandenen Messsystemen und zulässigen Grenzwerten der jeweiligen Sensoren. Die aus dieser Messung ermittelten maximalen Stellbereiche der Aktoren ergeben den zulässigen statischen Variationsraum (Stellbereich der Aktoren).

Neben der Bestimmung von statischen Variationsräumen, können ebenso dynamische Variationsräume ermittelt werden (siehe Anhang D). Letztere sind aber nach heutigem Stand nicht sinnvoll nutzbar zur Ableitung von Informationen für den folgenden Vermessungsprozess. Durch den exponentiell steigenden Aufwand ist die Bestimmung dessen bereits bei 2 Stellgrößen nicht mehr sinnvoll darstellbar.

In Bild 3.1 sind die Schritte bzw. die Struktur der Online-Vermessung aus den Bildern 1.5 und 2.3 vereinfacht dargestellt. Im Folgenden werden, nach einer Beschreibung der Grundlagen

(siehe Abschnitt 3.1), die Erweiterungen der Variationsraumbestimmung aus dem ersten Schritt zum Einsatz im Rahmen der Online-Vermessung vorgestellt. Hierfür werden ausgehend von geeigneten Grenzgrößen (siehe Abschnitt 3.2) ein geeigneter Messplan zur Bestimmung des Variationsraums entworfen (siehe Abschnitt 3.3). Mithilfe einer Grenzraumregelung kann dann eine Vermessung des Variationsraumes erfolgen (siehe Abschnitt 3.4) und so ausgehend von den gemessenen Daten eine einfache mathematische Beschreibung konvexer und konkaver Räume erfolgen (siehe Abschnitt 3.5).

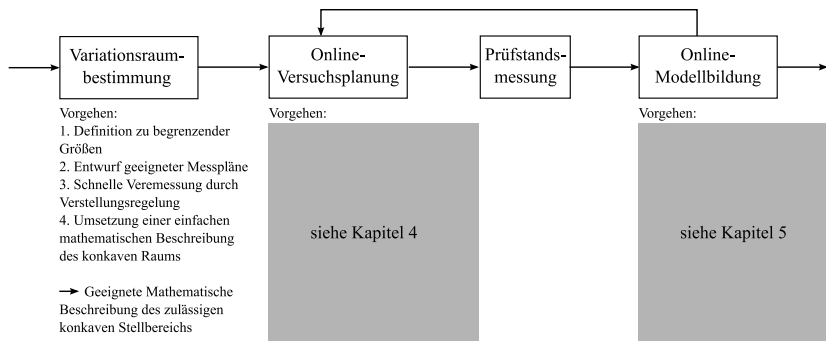


Bild 3.1: Vereinfachte Darstellung der einzelnen Schritte einer Vermessung und daraus resultierend das Vorgehen zum Entwurf einer durchgängigen Online-Methode im Rahmen dieser Arbeit bzw. in diesem Kapitel.

3.1 Grundlagen der Variationsraumbestimmung

Methoden zur Bestimmung von Variationsräumen sind arbeitspunktbasiert. Das heißt sie bestimmen für feste Einspritzmengen- und Drehzahl-Kombinationen jeweils einen Variationsraum. Die sich ergebende Anzahl an Variationsräumen entspricht dadurch der Anzahl an zuvor festgelegten Arbeitspunkten. Eine erste automatisiert ablaufende Methode stammt von AVL [50]. Dabei wird der gesamte Stellbereich sternförmig in vorgegebenen Schritten oder langsam kontinuierlich abgefahren bis die Grenzwerte erreicht werden. Zur schnellen Bestimmung des Variationsraums hat Daimler [127, 128, 3] die RHD-Methode entwickelt, bei der ein iteratives Verfahren eingesetzt wird, um den zulässigen konvexen Variationsraum in kurzer Zeit ausreichend genau bestimmen zu können. Dazu wird zunächst eine sternförmige Grundvermessung von einem sicheren Startpunkt aus durchgeführt. Aus diesen Punkten wird eine konvexe Hülle bestimmt. Ausgehend von den Mittelpunkten der Hyperflächen der Hülle wird die Vermessung entlang des Normalenvektors zu den ermittelten Variationsgrenzen iterativ fortgesetzt (siehe Bild 5.1). Zudem verwendet diese Methode nicht nur eine lokale (d.h. in der jeweiligen Vermessungsrichtung), sondern auch eine globale (d.h. für den gesamten Verstellbereich) Prädiktion, um die Messzeit, bei gleichzeitigem Gewinn an Informationen, zu reduzieren. Nach

Abschluss der Messung können alle zulässigen Stellgrößenkombinationen in Form einer konvexen Hülle dargestellt werden. Eine einschließende Hülle ist dabei konvex, wenn die gesamte Verbindungsstrecke zwischen je zwei Punkten innerhalb der Hülle ebenso wieder in der Hülle liegt. Ansonsten ist die Hülle konkav [22] (siehe Bild 3.2). Für deren Berechnung haben Barber und Co. den Quickhull-Algorithmus entwickelt [9]. Dieser hat den Vorteil, dass sowohl die Berechnung als auch die Auswertung sehr einfach ist und somit der Rechenaufwand gering ist.

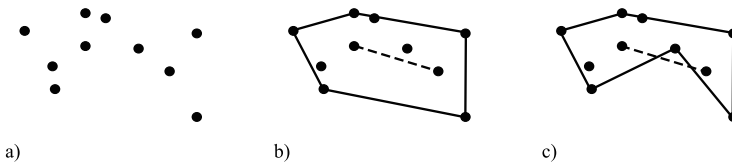


Bild 3.2: Vergleich einer konvexen und konkaven Hülle. a) Einzuhüllende Punktwolke b) Konvexe Hülle mit Verbindungsstrecke innerhalb der Hülle c) Konkave Hülle mit Verbindungsstrecke außerhalb der Hülle

In [85] wird eine Verstellstrategie zur Bestimmung des Variationsraums vorgestellt, welche die RHD-Methode (Rapid Hull Determination) speziell für den Fall konkaver Räume erweitert. Durch eine geeignete Auswahl des Startpunktes kann hierbei sowohl die Fahrbarkeit der Startpunkte als auch eine schnellere Bestimmung der Grenzwerte gewährleistet werden. Auch Yoshida [171] erweitert den RHD-Algorithmus um die Funktionalität, konkave Räume zu vermessen. Dabei wird basierend auf den in der ersten Iteration vermessenen Verstellwegen und der bestimmten konvexen Hülle der kürzeste Weg zum Mittelpunkt der Hyperfläche ausgewählt. Die CLA-Strategie (Continuous Limit Approach) [115] zur Motorvermessung stellt zur Bestimmung des Grenzraums die Aktoren kontinuierlich und integriert eine automatische Online-Reaktion auf Grenzverletzungen. Dabei erfolgt die Verstellung mit fester Geschwindigkeit bis eine Grenzverletzung auftritt. Daraufhin erfolgt ein Rückzug um 30% des gestellten Weges sowie eine Wiederannäherung bis kurz vor die Grenze. Diese Annäherungs- und Rückzugsstrategie wird bis zum Erreichen eines stabilen und sicheren Messpunktes in Grenznähe wiederholt. Der von der IAV entwickelte Advanced Boundary Finder (ABF) [59, 47] kombiniert die gleichzeitige multidimensionale Verstellung der Aktoren mit einer Anpassung der Schrittweiten. Wieder wird dabei eine Rückzugs- und Wiederannäherungsstrategie umgesetzt, um den tatsächlichen Grenzpunkt zu finden. Chaudoye [24] reduziert die Anzahl benötigter Messpunkte zur Bestimmung der konvexen Hülle, aufgespannt über den Ladedruck und die Luftmassenstrom, indem die physikalischen Grenzen als vier gerade Linien modelliert werden. Dadurch müssen nur noch zwei Punkte je Grenze gemessen werden, um den modellierten Raum vollständig aufspannen zu können. Neue Ansätze [151] integrieren die Bestimmung des zulässigen Variationsraums in die eigentliche Vermessung. Durch die online Identifikation eines Modells kann die iterative Messplanerstellung als modellbasierte Optimierungsaufgabe mit den zulässigen Stellbereichen als Nebenbedingungen formuliert werden. Dieses Verfahren setzt eine initiale Vermessung nicht nur innerhalb, sondern auch außerhalb des zulässigen Variationsraums

voraus. Das daraus entstehende initiale Modell stellt die Stellbereichsgrenzen in ausreichender Güte dar.

3.2 Grenzgrößen bei der Variationsraumbestimmung

Für eine automatisierte Echtzeit-Vermessung müssen alle begrenzten Ausgangsgrößen als Echtzeit-Messwert vorliegen. Messbare Größen, wie z.B. Emissionen, Temperaturen, Drücke oder der Schalldruckpegel können dabei direkt gemessen werden und der Automatisierung als Messgröße zur Verfügung gestellt werden. Weiterhin gibt es begrenzte Größen, welche nicht direkt gemessen werden können, wie z.B. Zündaussetzer. Diese müssen zur Berücksichtigung innerhalb der automatisierten Vermessung in Echtzeit berechenbar sein.

3.2.1 Arten der Grenzgrößen

Bei Grenzen wird zwischen harten und weichen Grenzen unterschieden [87]. Als harte Grenzen werden dabei Werte von Ausgangsgrößen bezeichnet, bei deren Überschreitung eine Beschädigung des Motors möglich ist. Diese dürfen unter keinen Umständen in einer automatisierten Vermessung überschritten werden. Im Gegensatz dazu gibt es weiche Grenzen, welche den Stellbereich der Applikation bestimmen. Diese Grenzwerte dürfen kurzfristig überschritten werden (z.B. Überspringen beim Einschwingvorgang einer dynamischen Anregung), aber ein dauerhafter Betrieb über diese Grenzwerte hinaus sollte vermieden werden. Je nach Art der Ausgangsgröße kann es sowohl eine harte als auch eine weiche Grenze geben oder aber ausschließlich eine weiche Grenze. Zu hohe Temperaturen, Zündaussetzer oder Zylinderspitzen drücke können eine Beschädigung des Motors zur Folge haben. Aus diesem Grund müssen diese im Rahmen einer automatisierten Vermessung hart begrenzt werden. Zusätzlich kann für diese Ausgangsgrößen eine weiche Grenze für die gewünschte Applikation definiert werden, innerhalb derer sich diese befinden dürfen. Dies ist z.B. der Fall bei Emissionen oder dem Motorgeräusch, die in der Regel nur weich begrenzt werden. Eine Überschreitung dieser Grenze hat keine motorschädlichen Folgen und folglich nur einen Einfluss auf die gewünschte Applikation. Durch Definition von einzuhaltenden Grenzwerten für NO_x , Partikel-Emissionen und den Schalldruckpegel können die maximal auftretenden Emissionen und Geräusche innerhalb einer Applikation beschränkt werden.

3.2.2 Bestimmung von Zündaussetzern als Grenzgröße

Die leistungsfähigste Methode zur Erkennung von Zündaussetzern ist der Einsatz von Zylinderdrucksensoren. Die Erkennung wäre auch über eine Analyse der Drehzahl möglich, jedoch stellt der Brennraumdruck unmittelbar direkte Informationen über den Verbrennungsverlauf zur Verfügung, während der Drehzahl die Schwingungen des Riemmentriebs als Störungen überlagert sind. Der Einsatz von Brennraumdrucksensoren stellt zum heutigen Zeitpunkt zwar eine kostenintensive Lösung dar, wird wegen der vielfältigen weiteren Einsatzbereiche jedoch an

Prüfständen sowie auch in der Serie genutzt. Mögliche Anwendungsgebiete sind dabei die Bestimmung des indizierten Motormoments, die zylinderindividuelle Schwerpunktlagenregelung oder auch die Zylindergleichstellung [60, 112, 168, 88].

Der in [168] entwickelte Frequenzbereichsansatz ermöglicht eine zylinderindividuellen Aussetzerdiagnose basierend auf einer globalen Aussetzererkennung. Dabei werden die Zylinderdrucksignale winkelsynchron tiefpassgefiltert und in Bereiche der Dauer eines Arbeitsspiels segmentiert. Die globale Erkennung stellt dabei basierend auf der Varianz σ_{seg}^2 des gefilterten und segmentierten Zylinderdruckverlaufs fest, ob in einem Segment ein Aussetzer aufgetreten ist. Anschließend wird anhand der individuellen Aussetzerdiagnose bestimmt, in welchem Zylinder welches Aussetzermuster aufgetreten sind. Aussetzer haben einen Anstieg der Varianz σ_{seg}^2 zur Folge. Durch die Vorgabe einer, von der Drehzahl und Last bedingten Schwelle σ_{lim}^2 kann abhängig von der aktuellen Varianz ein Aussetzer erkannt werden. Die Schwelle wurde in [168] als Polynom dritter Ordnung definiert, wobei deren Parameter vorneweg mittels der Least-Squares Schätzung aus der Stationärvermessung von 59 Motorarbeitspunkten bestimmt worden sind.

Um Zündaussetzer als Grenze im Rahmen einer automatischen Vermessung zu nutzen, wird anstelle eines modellgestützten Verfahrens ein einfaches signalbasiertes Verfahren eingesetzt. Vorteil ist hierbei die einfache Umsetzung und schnelle Berechnung sowie die Vermeidung von Vorabmessungen. Dadurch ist die Lokalisierung der Zylinder sowie die Unterscheidung von verschiedenen Aussetzerkombinationen nicht möglich. Da die Zündaussetzer in diesem Fall jedoch lediglich als Grenze verwendet werden, spielt dies keine Rolle. Sobald ein Aussetzer erkannt wird, ist die Grenze des Stellbereichs erreicht und ein weiteres Stellen der Aktoren unzulässig, unabhängig davon in welchem Zylinder der Aussetzer aufgetreten ist.

Das in [88] entwickelte Druckindiziersystem stellt verschiedene, aus dem Brennraumdruckverlauf extrahierte Merkmale in Echtzeit zur Verfügung. Der Fachbegriff Indizierung bezeichnet dabei das schnelle Messen des Brennraumdrucks p_{Zyl} über die Kurbelwellenwinkelstellung φ_{KW} . Der Druck im Brennraum wird dabei über einen piezoelektrischen Quarzsensormessung [29]. Durch den Druck auf eine Membran entsteht eine Ladung Q_s im Sensor, welche über einen Ladungsverstärker in ein Spannungssignal U_s umgewandelt wird. Dieses Spannungssignal wird dann über eine AD-Karte dem Echtzeit-Indiziersystem zur Verfügung gestellt. Die AD-Karte wird dabei mit einem von einer Timerkarte aufbereiteten Winkelsignal $\tilde{\varphi}$ getriggert, um eine kurbelwellensynchrone Aufzeichnung des Zylinderdrucks zu ermöglichen [88]. Dieses wird dann aufbereitet, indem eine Filterung, eine Einstellung auf den oberen Totpunkt (OT) sowie eine Nulllinien- und Driftkorrektur erfolgt. Ausgehend von den aufbereiteten Drucksignalen können verschiedene Merkmale und Kenngrößen aus dem Druckverlauf bestimmt werden.

Die umgesetzte Erkennung von Zündaussetzern basiert auf der Kenngröße des maximalen Differenzdrucks $\Delta p_{\text{Zyl,max}}$. Dieser ist im allgemeinen ein Maß für die Laufunruhe des Motors, welche durch Zündaussetzer maßgeblich beeinflusst wird. Die Bestimmung von $\Delta p_{\text{Zyl,max}}$ erfolgt aus der Differenz des gemessenen Drucks $p_{\text{Zyl}}(\varphi)$ und des rekonstruierten Schleppdrucks

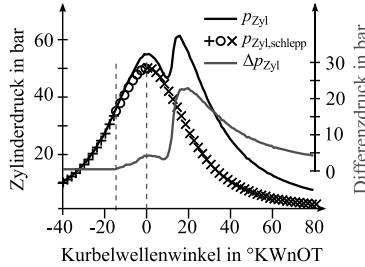


Bild 3.3: Verlauf des gemessenen Zylinderdrucks p_{Zyl} , des rekonstruierten Schleppdrucks $p_{Zyl,schlepp}$ sowie des sich ergebenden Differenzdrucks Δp_{Zyl} .

$p_{Zyl,schlepp}(\varphi)$:

$$\Delta p_{Zyl,max} = \max_{\varphi \in [-360^\circ \text{KWnOT}, 360^\circ \text{KWnOT}]} (p_{Zyl}(\varphi) - p_{Zyl,schlepp}(\varphi)). \quad (3.1)$$

Dabei muss der Schleppdruck, der im ungefeuerten Betrieb entstehen würde, rekonstruiert werden. Hierfür wird bis zum frühesten Brennbeginn φ_{VB} (i.d.R. auf -15°KWnOT abgeschätzt [97, 88]) der gemessene Druck verwendet. Zwischen dem Brennbeginn und dem oberen Totpunkt wird der Druck über die Gleichung der polytropen Kompression berechnet. Ab dem oberen Totpunkt kann der restliche Druckverlauf durch Spiegelung des Drucks am OT bestimmt werden. Der Zylinderdruckverlauf folgt nach [114, 88] aus

$$p_{Zyl,schlepp}(\varphi) = \begin{cases} p_{Zyl}(\varphi) & \text{für } \varphi = (-360^\circ \text{KWnOT}, \varphi_{VB}] \\ p_{Zyl}(\varphi_{VB}) \cdot \left(\frac{V(\varphi_{VB})}{V(\varphi)} \right)^n & \text{für } \varphi = (\varphi_{VB}, \varphi_{OT}] \\ p_{Zyl,schlepp}(-\varphi) & \text{für } \varphi = (\varphi_{OT}, 360^\circ \text{KWnOT}) \end{cases} \quad (3.2)$$

mit dem vom Kurbelwinkel abhängigen Zylindervolumen

$$V(\varphi) = rA \left((1 - \cos(\varphi)) + \frac{l}{r} \left(1 - \sqrt{1 - \left(\frac{r}{l} \sin(\varphi) \right)^2} \right) \right) + \frac{2rA}{\varepsilon - 1}. \quad (3.3)$$

Dabei sind l die Länge der Pleuelstange, r der Radius der Kurbelwelle, A die Kolbenfläche, n der Polytropenexponent und ε das Verdichtungsverhältnis. In Bild 3.3 ist für einen gemessenen Druckverlauf im Arbeitspunkt ($1500 \text{ min}^{-1} / 20 \frac{\text{mm}^3}{\text{Hub}}$) der sich ergebende rekonstruierte Schleppdruckverlauf sowie Differenzdruckverlauf dargestellt. Beim Winkel 20°KWnOT ergibt sich der gesuchte maximale Differenzdruck $\Delta p_{Zyl,max}$.

Die maximalen Zylinderdruckdifferenzen jedes Zylinders werden vom Indiziersystem in Echtzeit der Vermessung als aus Messdaten berechnete Kenngröße zur Verfügung gestellt. Die Beurteilung dieser zur Zündaussetzererkennung erfolgt über statistische Bewertungsmaße. Zum Einsatz kommen dabei der arithmetische Mittelwert

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x(i), \quad (3.4)$$

welcher dem Erwartungswert bei ergodischen Prozessen entspricht (ergodisch $\hat{=}$ Varianz des arithmetischen Mittelwerts konvergiert für $n \rightarrow \infty$ gegen Null) sowie die Varianz

$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x(i) - \mu_x)^2, \quad (3.5)$$

als Maß für die Streuung bzw. die mittlere quadratische Abweichung vom Mittelwert und die Standardabweichung

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x(i) - \mu_x)^2}. \quad (3.6)$$

Zur Erkennung von Zündaussetzern wird die Standardabweichung des maximalen Differenzdrucks über alle N Zylinder des Versuchsträgers bestimmt:

$$\mu_{\Delta p_{\text{Zyl,max}}} = \frac{1}{N} \sum_{i=1}^N \Delta p_{\text{Zyli,max}} \quad (3.7)$$

$$\sigma_{\Delta p_{\text{Zyl,max}}} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\Delta p_{\text{Zyli,max}} - \mu_{\Delta p_{\text{Zyl,max}}})^2} \quad (3.8)$$

Dabei wird die Erkennungsschwelle unabhängig vom Arbeitspunkt definiert, indem diese auf die initiale Standardabweichung der Zylinder $\sigma_{\Delta p_{\text{Zyl,init}}}$ im sicheren Zentralpunkt bei einer lokalen Vermessung bezogen wird. Dadurch kann die Schwelle direkt als Grenze des Konfidenzintervalls der Standardabweichung vorgegeben werden. Mit Vorgabe einer Schwelle von $4 \cdot \sigma_{\Delta p_{\text{Zyl,init}}}$, bei der (unter Annahme einer Normalverteilung) über 99,99 % aller Messwerte innerhalb des Intervalls sind, folgt die Erkennung eines Zündaussetzers gemäß der folgenden Fallunterscheidung:

$$v_{\text{Aussetzer}} = \begin{cases} 0 & \text{für } \sigma_{\Delta p_{\text{Zyl,max}}} < 4 \cdot \sigma_{\Delta p_{\text{Zyl,init}}} \\ 1 & \text{für } \sigma_{\Delta p_{\text{Zyl,max}}} \geq 4 \cdot \sigma_{\Delta p_{\text{Zyl,init}}} \end{cases} \quad (3.9)$$

Das Auftreten eines Zündaussetzers ist im Gegensatz zu anderen Grenzen ein diskretes Ereignis. Um die Robustheit gegenüber einzelnen Ausreißern bei der Zylinderdruckmessung sowie einzelnen Signalen außerhalb des Konfidenzintervalls zu erhöhen, wird zusätzlich ein Aussetzerzähler integriert. Dieser zählt die auftretenden Aussetzer aller Zylinder bis zu einer vorgegebenen Grenze $s_{\text{Aussetzer}}$, bevor eine Grenzverletzung erkannt wird. Wird kein Aussetzer mehr erkannt, bevor der Grenzwert erreicht ist, zählt der Zähler wieder rückwärts, jedoch mit der halben Schrittweite. Daraus folgt, dass der Zähler wieder zurückgesetzt werden kann, indem bei mindestens der doppelten Anzahl an Zündtakten kein Aussetzer vorliegt. Durch die Asymmetrie wird die Gutprüfung der Diagnose im Anschluss an einen einzelnen Zündaussetzer robust im Vergleich zu der Schlechtprüfung der Diagnose.

In Bild 3.4 ist das Vorgehen zusammenfassend dargestellt. Innerhalb der ersten Zündtakte werden zwei Zündaussetzer erkannt, aber keine dauerhafte Laufunruhe. Demzufolge wird die vorgegebene Grenze $s_{\text{Aussetzer}} = 4$ nicht erreicht und nach 4 weiteren Zündtakten der Zähler zurückgesetzt. Im weiteren Verlauf treten vermehrt Aussetzer auf, so dass zum 14. Zündtakt die Grenze der Motorlaufruhe bzw. der Zündaussetzer erreicht wird.

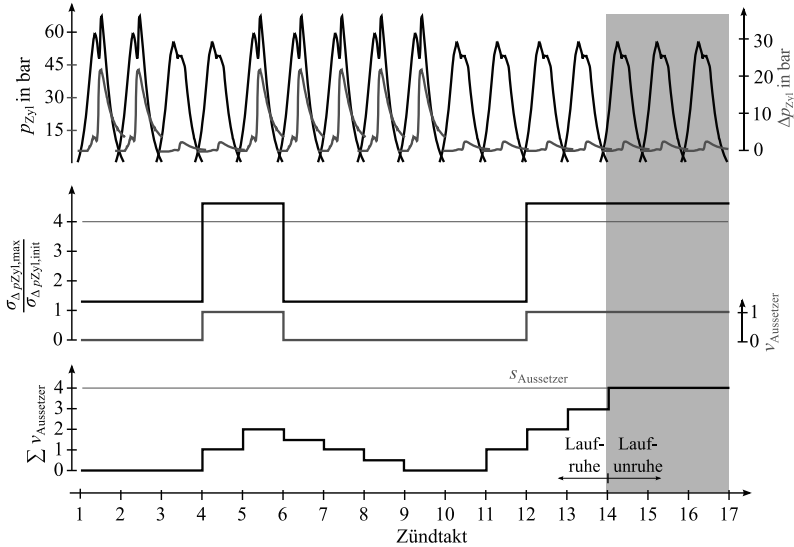


Bild 3.4: Bestimmung der Grenze der Laufunruhe bzw. Zündaussetzer. Die Erkennung eines Zündaussetzers $v_{\text{Aussetzer}}$ erfolgt, indem die Standardabweichung des maximalen, gemessenen Differenzdrucks über alle 4 Zylinder die Schwelle $4 \cdot \sigma_{\Delta p_{\text{Zyl,lim}}}$ überschreitet. Sobald eine vordefinierte maximale Anzahl an Zündaussetzern $s_{\text{Aussetzer}}$ aufgetreten ist, wird eine Grenzverletzung detektiert.

3.3 Messpläne zur Bestimmung von Variationsräumen

Für die Bestimmung von Variationsräumen können verschiedenen Methoden der Messplanerstellung zum Einsatz kommen, welche einen zu vermessenden Kandidatensatz erstellen.

Vollständiger Raster-Messplan

Eine vollständig rasterförmige Abdeckung des Versuchsraumes, bei der eine Rasterung sowohl zwischen den minimalen als auch den maximalen Werten jeder Stellgröße durchgeführt wird, ist aufgrund des *Fluchs der Dimensionen* nur bei kleiner Dimension P und wenigen Rasterstufen L sinnvoll. Die Anzahl der Messpunkte N erhöht sich exponentiell gemäß

$$N = L^P. \quad (3.10)$$

Zentral zusammengesetzter Messplan

Eine weitere Möglichkeit besteht im Einsatz zentral zusammengesetzter Messpläne (CCD, central composite designs). Dabei wird bei P Stellgrößen ein würfelförmiger Messplan mit $2P + 1$

Messpunkten mit einem sternförmigen Messplan mit $2 \cdot P + 1$ Messpunkten kombiniert, indem die Mittelpunkte aufeinander gelegt werden. Nachteilig ist, dass hierbei nur eine grobe Erfassung der Hülle erfolgt. Zur genaueren Bestimmung des zulässigen Variationsraums muss die Anzahl der Messplanpunkte an den Rändern des mehrdimensionalen Raumes erhöht werden.

Hüllraster-Messplan

Ein quaderförmiger Hüllraster-Messplan ähnelt dem eines vollständigen Rasters. Im Gegensatz zur rasterförmigen Abdeckung des Versuchsraums werden dabei aber Stellgrößenkombinationen \mathbf{u} , bei denen alle Stellgrößen nicht gleichzeitig am Versuchsraumrand sind, ignoriert. Für \mathbf{u} gilt entsprechend:

$$\mathbf{u}_{\min} \leq \mathbf{u} \leq \mathbf{u}_{\max} \quad (3.11)$$

Quaderförmige Hüllraster-Messpläne eignen sich zur Echtzeit-Vermessung des Variationsraums in Kombination mit einer rampenförmigen Verstellung der Aktoren. Dabei werden nur die extremalen (maximalen und minimalen) Stellgrößenamplituden als zu vermessende Kombination generiert, welche somit ausschließlich am Rand des Variationsraums liegen. Die Messpunktanzahl reduziert sich im Vergleich zu einem vollständigen Raster um $(L - 2)^P$ zu:

$$N = L^P - (L - 2)^P \quad (3.12)$$

Die Anzahl der Stellgrößen P in Kombination mit der Anzahl an Rasterstufen L , welche der gewünschten Genauigkeit entsprechen, definieren dabei direkt die Anzahl der Hüllrasterpunkte und somit die zu vermessende Anzahl an Messpunkten. In Bild 3.5 sind ein zentral zusammengesetzter, ein Raster- und Hüllraster-Messplan mit $L = 7$ für zwei Stellgrößen u_1 und u_2 dargestellt. Durch die Erhöhung der Anzahl an Messpunkten von 8 beim CCD auf 49 Messpunkte beim Raster kann der sich ergebende Variationsraum genauer beschrieben werden. Beim Hüllraster-Messplan werden im Vergleich zum Raster-Messplan nur noch 24 Messpunkte benötigt, um die gleiche Genauigkeit des zulässigen Variationsraums zu erhalten. Es werden nur Punkte generiert, welche ausschließlich am Rand des Variationsraums liegen und folglich einen direkten Einfluss auf die Genauigkeit der Bestimmung des zu vermessenden Variationsraums haben. Weiterhin ist der Hüllraster-Messplans unabhängig von der Messpunktreihenfolge, welche bei sowohl dem CCD als auch dem Raster-Messplan einen großen Einfluss auf den bestimmten Variationsraum hat.

3.4 Schnelle Bestimmung des statischen Variationsraums

Eine Beschleunigung der Variationsraumbestimmung kann durch die Kombination einer kontinuierlichen adaptiven Verstellung der Aktoren mit einer Online-Reaktion auf die vorgegebenen Grenzen erfolgen. Dadurch kann der Zeitaufwand, welcher sich durch eine schrittweise Annäherung, die Annäherungs- und Rückzugsstrategie sowie die Einschwingzeit zum Erreichen der stationären Zustände ergibt, reduziert werden. Das Blockschaltbild des entwickelten CHA-Verfahrens (Controlled Hull Approach) ist in Bild 3.6 für eine Stell- und eine Ausgangsgröße dargestellt.

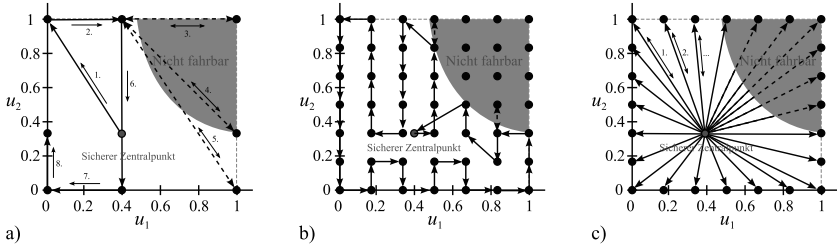


Bild 3.5: Vergleich verschiedener Messpläne zur Bestimmung des Variationsraumes. a) Zentral zusammengesetzter Messplan mit 8 Messpunkten, b) Raster-Messplan mit 49 Messpunkten und c) Hüllraster-Messplan mit 24 Messpunkten

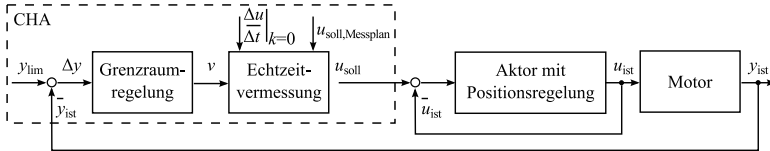


Bild 3.6: Blockschaltbild der Grenzraumregelung zur schnellen Bestimmung des Variationsraums

Mit dem Abstand der online gemessenen Ausgangsgröße y_{ist} zu ihrem vorgegebenen Grenzwert y_{lim} erfolgt eine Grenzraumregelung. Die Stellgröße der Regelung ist dabei der Adaptionfaktor v der Verstellgeschwindigkeit des Aktors. Zusammen mit einer vorgegebenen Anfangsgeschwindigkeit $\left. \frac{\Delta u}{\Delta t} \right|_{k=0}$ und der Änderung der Sollgröße

$$\Delta u = u_{\text{soll,Messplan}} - u_{\text{soll}}(k-1) \quad (3.13)$$

wird die aktuelle Verstellgeschwindigkeit für die Echtzeit-Vermessung bestimmt:

$$\left. \frac{\Delta u}{\Delta t} \right|_{k \neq 0} = \begin{cases} v \cdot \left. \frac{\Delta u}{\Delta t} \right|_{k=0} & \text{für } y_{\text{ist}} < y_{\text{lim}} \\ 0 & \text{für } y_{\text{ist}} = y_{\text{lim}} \\ -v \cdot \left. \frac{\Delta u}{\Delta t} \right|_{k=0} & \text{für } y_{\text{ist}} > y_{\text{lim}} \end{cases} \quad (3.14)$$

Die Ausgangsgröße der Echtzeit-Vermessung folgt daraus zu

$$u_{\text{soll}}(k) = u_{\text{soll}}(k-1) + T_0 \cdot \left. \frac{\Delta u}{\Delta t} \right|_{k \neq 0} \quad (3.15)$$

Der Grenzraumregler regelt somit die Verstellgeschwindigkeit des Aktors in Abhängigkeit von der Entfernung der Ausgangsgröße zur Grenze, wobei die Geschwindigkeit von der Echtzeit-Vermessung in einen entsprechenden zeitlichen Verlauf der Stellgröße gewandelt wird.

Bei der Regelung wird die Verstellgeschwindigkeit umso kleiner, je näher y_{ist} an die Grenze y_{lim} vorrückt. Bei der Adaption der Geschwindigkeit in Richtung schnellerer Werte ist zu beachten,

dass die sich aus der adaptierten Geschwindigkeit und dem Abstand zwischen zwei Messplanpunkten ergebende Anstiegszeit T_R nicht ein vorgegebenes Minimum entsprechend [70]

$$T_{R,\min} = \left| \frac{u_{\text{soll,Messplan}} - z}{v \cdot \left. \frac{\Delta u}{\Delta t} \right|_{k=0}} \right| \gg \frac{0.5}{\omega_{\max}} \quad (3.16)$$

$$\rightarrow |v| < \left| \frac{\omega_{\max} \cdot (u_{\text{soll,Messplan}} - z)}{0.5 \cdot \left. \frac{\Delta u}{\Delta t} \right|_{k=0}} \right|$$

unterschreitet, um die dynamische Anregung des Prozesses gering zu halten. Dabei stellen ω_{\max} die höchste für die Identifikation relevante Frequenz und z den sicheren Zentralpunkt dar. Eine zu schnelle Rampe könnte ein starkes dynamisches Überspringen der Ausgangsgröße bis über den harten Grenzwert hinaus zur Folge haben.

Die Echtzeit-Vermessung stellt mithilfe der geregelten Geschwindigkeit den Sollwert des Aktors u_{soll} kontinuierlich in Richtung des Hüllrasterpunkts $u_{\text{soll,Messplan}}$. Letzter wird aus dem Messplan bestimmt. Bei der Verstellung in Richtung des Hüllrasterpunkts muss dabei zwischen zwei Fällen unterschieden werden:

1. Sollwert $u_{\text{soll,Messplan}}$ aus dem Messplan wird erreicht, da kein Grenzwert überschritten wird: $y_{\text{ist}} \leq y_{\text{lim}}, u_{\text{soll}} = u_{\text{soll,Messplan}}$.
2. Sollwert $u_{\text{soll,Messplan}}$ aus dem Messplan wird nicht erreicht, da ein Grenzwert überschritten wird: $y_{\text{ist}} \geq y_{\text{lim}}, u_{\text{soll}} \neq u_{\text{soll,Messplan}}$.

Im ersten Fall erreicht der Aktor seinen maximalen Stellwert. Da keine Grenzverletzung an dieser Stelle auftritt, wird der Wert der Stellgröße an dieser Stelle über eine definierte Zeitdauer gemittelt und abgespeichert, bevor im Anschluss die Echtzeit-Vermessung den Aktor sprunghaft zurück zum sicheren Zentralpunkt z stellt. Von dort aus wird die Vermessung des nächsten Hüllrasterpunkts gestartet.

In zweiten Fall wird online eine Grenzverletzung auf dem Weg zum Hüllrasterpunkt erkannt. Die Regelung des Grenzraums erfolgt gemäß Gleichung (3.14), indem sie bei Erkennung des Grenzwertes ($y_{\text{ist}} = y_{\text{lim}}$) der Ausgangsgröße die Verstellung anhält bzw. bei einem Überschreiten des Grenzwertes ($y_{\text{ist}} > y_{\text{lim}}$) automatisch rückwärts verstellt. Sobald die Verstellgeschwindigkeit null ist, d.h. die Ausgangsgröße sich auf der Grenze befindet, hat der Aktor seinen maximal zulässigen Stellwert erreicht. Der Wert der Stellgröße u_{ist} wird über eine definierte Zeit gemittelt und nach Ablauf dieser abgespeichert. Im Anschluss stellt die Echtzeit-Vermessung den Aktor wieder in den sicheren Zentralpunkt z und fährt mit dem nächsten Hüllrasterpunkt fort.

In Bild 3.7 ist das CHA-Verfahren im Vergleich zum CLA-Verfahren (siehe Abschnitt 3.1 bzw. [115]) für die Vermessung mit einer Stell- und einer Ausgangsgröße dargestellt. Der Prozess, der zum Vergleich verwendet wurde, ist ein PT_1 System mit Verstärkung von $K = 100$, Zeitkonstante $T_1 = 0.2 \text{ s}$ und normierter Eingangsgröße $0 \leq u_{\text{soll}} \leq 1$. Als sicherer Zentralpunkt wird $u_{\text{soll}} = 0.5$, als Grenzwerte der simulierten Ausgangsgröße $y_{\text{lim,min}} = 10$ und

$y_{\text{lim,max}} = 80$ vorgegeben. Das sich zur Bestimmung des Variationsraums ergebende Hüllraster besteht in diesem eindimensionalen Fall aus zwei Punkten, $u_{\text{sol}} = 0$ sowie $u_{\text{sol}} = 1$. Zusätzlich wurde der Messplan um einen weiteren Messpunkt ergänzt, welcher innerhalb des zulässigen Raums liegt, um den Unterschied zwischen dem Ansteuern von Hüllpunkten und nicht-Hüllpunkten darzustellen. Die einzelnen Punkte werden bei CHA-Verfahren ausgehend von dem sicheren Zentralpunkt vermessen, während beim CLA-Verfahren der vorherige Messpunkt als Startpunkt dient. Als initiale Verstellgeschwindigkeit wird für beide Verfahren $\frac{\Delta u}{\Delta t} = 0.1 \frac{1}{s}$ vorgegeben.

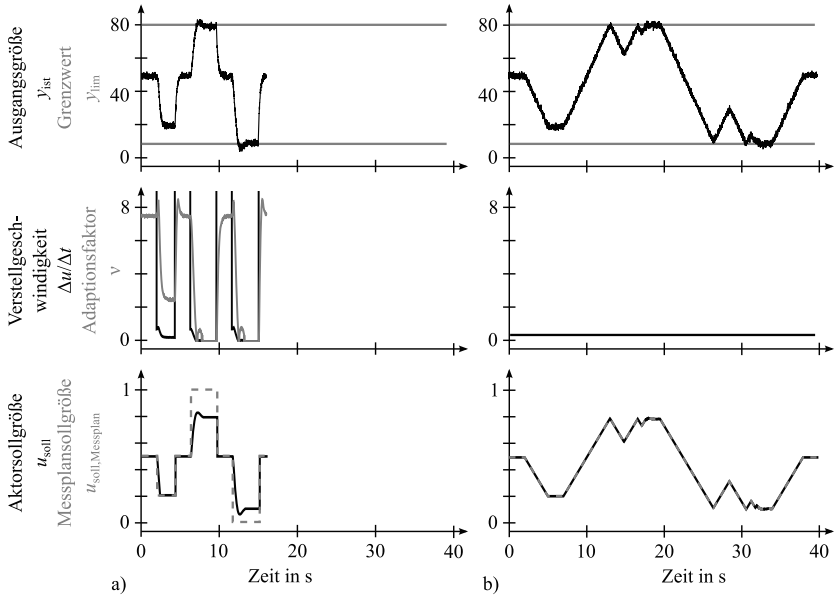


Bild 3.7: Vergleich der Verfahren CHA und CLA für die Variationsraumbestimmung eines simulierten PT_1 Systems mit vorgegebenen Grenzen bei $y_{\text{lim,min}} = 10$ und $y_{\text{lim,max}} = 80$. In a) sind für das CHA-Verfahren der zeitliche Verlauf der Ausgangsgröße y_{ist} , der Verstellgeschwindigkeit des Aktors $\frac{\Delta u}{\Delta t}$, des Adaptionsfaktors ν sowie der Aktorsollgröße u_{sol} und Messplansollgröße $u_{\text{sol,Messplan}}$. In b) die entsprechenden Verläufe für das CLA-Verfahren, wobei dieses keinen Adaptionsfaktor hat. Durch die Regelung der Verstellgeschwindigkeit basierend auf dem Abstand zum Grenzwert der Ausgangsgrößen erfolgt die Verstellung der Sollgrößen und dadurch die Bestimmung der statischen Hüllraster-Messpunkte mit dem CHA-Verfahren schneller als bei bisherigen Verfahren.

Die sich ergebenden zeitlichen Verläufe der Ausgangsgröße y_{ist} und der Grenze y_{lim} , der sich daraus ergebende geregelte Adaptionsfaktor ν sowie die daraus folgende Verstellgeschwindigkeit $\frac{\Delta u}{\Delta t}$ und Aktorsollgröße u_{sol} bzw. $u_{\text{sol,Messplan}}$ sind für die beiden Verfahren in Bild 3.7 für das Beispielsystem dargestellt. Es ist erkennbar, dass beide Methoden geeignet sind, die Ak-

toren bei Erreichen der vorgegebenen Grenzwerte der Ausgangsgröße anzuhalten. Im Fall des CHA-Verfahrens erfolgt die Regelung der Verstellgeschwindigkeit kontinuierlich auf Basis des erkannten Abstandes zum Grenzwert y_{lim} . Das CLA-Verfahren verstellt dagegen die Stellgröße mit einer konstanten Geschwindigkeit $\frac{\Delta u}{\Delta t} = 0.1 \frac{1}{s}$. Während für den ersten Messpunkt, welcher innerhalb des zulässigen Stellbereichs liegt, die Verstellung bei beiden Verfahren angehalten und abgeschlossen wird, erfolgt für das Verstellen auf einen Messpunkt am Rand des zulässigen Stellbereichs eine Nachregelung aufgrund des leichten Überschwingens. Das CLA-Verfahren regelt dabei den Sollwert direkt auf den Grenzwert der Ausgangsgröße ein, während das CHA-Verfahren eine Rückzugstrategie bei Erreichen eines Grenzwertes anwendet. Dadurch kann im Fall des Beispielsystems eine Zeitersparnis von ca. 58 % (15 s im Vergleich zu 35 s) für die Vermessung mit dem CHA-Verfahren erreicht werden. Die Vermessung einer einzelnen Grenze dauert im Fall des CHA-Verfahrens 6.5 s, während das CLA-Verfahren 14.1 s benötigt, was einer Zeitersparnis von ca. 54 % entspricht.

Das CHA-Verfahren lässt sich ebenfalls für die Vermessung von MIMO-Prozessen (Multiple Input Multiple Output), mit mehreren Ein- und Ausgangsgrößen verwenden. Dabei ist zu berücksichtigen, dass sowohl mehrere Stellglieder als auch verschiedene zu begrenzende Ausgangsgrößen vorhanden sind. Um die Grenzraumregelung für N begrenzte Ausgangsgrößen zu erweitern, wird für jede eine separate Grenzraumregelung umgesetzt. Aus den sich so ergebenden N Adaptionfaktoren wird der aktuell relevante bestimmt:

$$v_{\min} = \min(v_1, v_2, \dots, v_N) \quad (3.17)$$

Die Erweiterung für mehrere Stellgrößen erfolgt durch Anpassung der Verstellgeschwindigkeiten aller Aktoren mit dem bestimmenden Adaptionfaktor v_{\min} .

In Bild 3.8 ist das 2-dimensionale Hüllraster mit der sich ergebenden Hülle des zulässigen Variationsraums sowie der zeitliche Verlauf der Ein- und Ausgangsgrößen einer Variationsraumvermessung mit den zwei Stellgrößen u_{HDAGR} (PWM-Signal des Hochdruck-Abgasrückführventils) und u_{VTG} (PWM-Signal der variablen Leitschaufeln eines Turboladers) dargestellt. Ausgehend von dem sicheren Zentralpunkt bei $u_{\text{HDAGR}} = u_{\text{VTG}} = 60\%$ werden nacheinander alle Punkte des Hüllrasters angefahren, mit aktivierter Regelung der Verstellgeschwindigkeit $\frac{\Delta u_{\text{HDAGR}}}{\Delta t}$ bzw. $\frac{\Delta u_{\text{VTG}}}{\Delta t}$ basierend auf dem Abstand Ausgangsgröße NO_x zu ihrer Grenze. Die Vermessung erfolgte mit den vorgegebenen Grenzen $100 \text{ ppm} \leq \text{NO}_x \leq 700 \text{ ppm}$ sowie unter der Randbedingung, dass keine Zündaussetzer auftreten. Aus dem Zeitverlauf der Ausgangsgrößen wird deutlich, dass die Grenzraumregelung die Ausgangsgrößen immer auf ihre Grenzwerte einregelt. Die Vermessung des Variationsraums dauerte bei 16 Hüllrasterpunkten insgesamt 508 s.

Durch den Einsatz des CHA-Verfahrens kann eine schnelle Bestimmung des Variationsraums erreicht werden, indem eine kontinuierliche sowie späte Online-Adaption der Verstellgeschwindigkeit im laufenden Betrieb erfolgt. Die Zeitersparnis ist dabei von den Eingangs- und Ausgangsgrößen sowie von der gewünschten Genauigkeit des Variationsraumes abhängig und steigt proportional mit der Anzahl an zu vermessenden Hüllrasterpunkten.

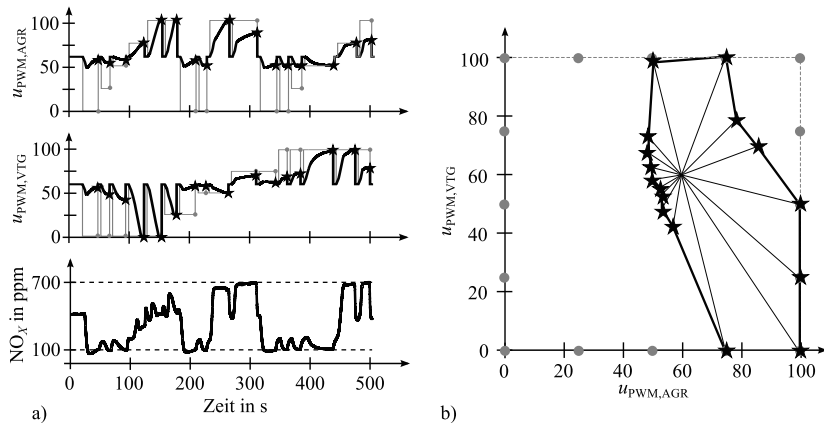


Bild 3.8: Zeitlicher Verlauf der Ein- und Ausgangsgrößen sowie der sich ergebende Variationsraum einer Variationsraumvermessung mit den Stellgrößen u_{HDAGR} und u_{VTG} sowie den Grenzen $100 \text{ ppm} \leq NO_x \leq 700 \text{ ppm}$ und keine Zündaussetzer.

3.5 Ansätze zur einfachen Beschreibung von statischen konkaven Variationsräumen

Die Beschreibung von Variationsräumen als konvexe Hülle ist nur eingeschränkt gültig und nicht universell einsetzbar. Sowohl in hochdimensionalen Variationsräumen als auch verstärkt beim Ottomotor sind vermehrt Bereiche als fahrbar durch eine konvexe Hülle beschrieben, obwohl sie nicht zulässig sind. In Bild 3.9 ist für die Verstellung der Last und des Zündzeitpunktes beim Ottomotor der sich durch die Klopfgrenze ergebende zulässige Variationsraum nach [85] dargestellt. Die sich nach erfolgter Vermessung ergebende konvexe Hülle ist in Bild 3.9 a) dargestellt, die konkave Hülle in b). Die Auswertung zeigt, dass die Beschreibung des zulässigen Variationsraums mit einer konvexen Hülle nicht ausreichend ist, da hierdurch Stellgrößen automatisiert in Bereichen gestellt würden, die ein Klopfen aufweisen.

Die Beschreibung des konkaven Variationsraums kann mithilfe von Splines erfolgen. Letztere haben jedoch den Nachteil, dass zur Interpolation die Daten äquidistant verteilt und monoton sein müssen, was jedoch im Rahmen einer Variationsraumvermessung nicht gegeben ist. Ansätze zur komplexeren Darstellung des Variationsraumes, wie z.B. Mithilfe von Support Vector Machines [13] oder RBF-Netzen [123, 83], sind Stand der aktuellen Forschung. Ein deutlicher Nachteil ist jedoch, dass sie viele stationäre Messungen benötigen, bis eine ausreichende Genauigkeit erreicht ist. Dies bedingt eine erhöhte Messzeit und ist somit für den Einsatz im Rahmen einer Online-Methode weniger geeignet.

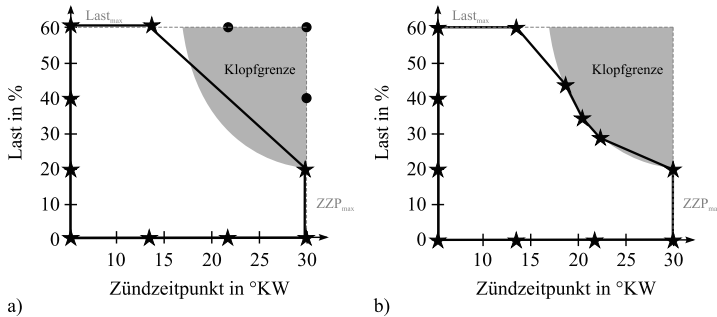


Bild 3.9: Darstellung des Variationsraumes von Last und Zündzeitpunkt für die Klopfgrenze beim Ottomotor [85] als a) konvexe und als b) konkave Hülle.

3.5.1 Einfache konkave Variationsraumdarstellung

Die aus der Variationsraumbestimmung erhaltenen, maximal zulässigen Stellgrößenkombinationen werden üblicherweise mit Hilfe der Delaunay-Triangulation [4] und dem Quickhull-Algorithmus [124, 9] in Form einer konvexen Hülle dargestellt.

Delaunay-Triangulation

Die Delaunay-Triangulation ist die Verkettung einer Punktemenge in ein Simplexnetz. Ein Simplex ist dabei die Verallgemeinerung eines Dreiecks aus dem Zweidimensionalen bzw. einer Pyramide aus dem Dreidimensionalen in höhere Dimensionen. Mithilfe der Triangulation kann die von der Punktemenge abgedeckte Fläche beschrieben werden. Vorteil gegenüber der gitterförmigen Verteilung eines Rasternetzes ist, dass die Stützstellen beliebig im Raum verteilt sein können. Die Delaunay-Triangulierung generiert im Vergleich zu anderen Verfahren kompakte Simplexe [165]. Diese werden dabei so generiert, dass die Hyperkugel eines jeden Simplex keine weitere Stützstelle von Nachbarsimplexen enthält (siehe Bild 3.10).

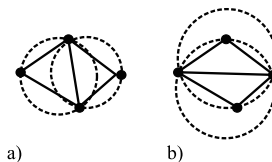


Bild 3.10: Vergleich einer a) Delaunay-Triangulation und einer b) Nicht-Delaunay-Triangulation (Umkreise enthalten Stützstellen von Nachbardreiecken) [114].

Quickhull-Algorithmus

Der Quickhull-Algorithmus ist ein effizienter Algorithmus zur Berechnung der konvexen Hülle einer Punktemenge. Eine konvexe Hülle entspricht der Verbindung aller Extrempunkte unter Einschließung aller Mengenpunkte. Zur Bestimmung dieser Stützstellen der Hülle wird die Punktemenge in immer kleinere Teilmengen rekursiv partitioniert, bis nur noch der Start- und Endpunkt der Trenngerade Bestandteil der Teilmenge sind. Die so bestimmte Gerade ist dann ein Segment der einschließenden konvexen Hülle. Das Vorgehen bei dem Quickhull-Algorithmus ist in Bild 3.11 zusammenfassend dargestellt.

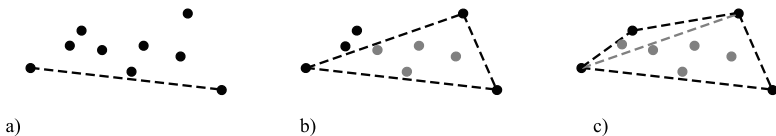


Bild 3.11: Vorgehen des Quickhull-Algorithmus zur Bestimmung einer die Punktemenge einschließenden konvexen Hülle.

Ausgehend von der Startgeraden zwischen den zwei horizontalen Extrempunkten (Bild 3.11 a)) wird mit dem am weitesten zur Geraden entfernten Punkt ein Dreieck gebildet (Bild 3.11 b)). Alle Punkte innerhalb des neu entstandenen Dreiecks können nicht mehr hüllenbildende Punkte sein. Anschließend wird wieder mit dem entferntesten Punkt ein neues Dreieck gebildet. Die inneren Punkte sind wieder keine hüllenbildende Punkte und werden ignoriert. Zusätzlich wird die ursprüngliche Seite des neuen Dreiecks ignoriert (Bild 3.11 c)). Dieses Vorgehen wird rekursiv solange fortgeführt, bis keine freien Punkte mehr vorhanden sind. Der gefundene Polygonzug ist die konvexe Hülle der Punktemenge.

Der große Vorteil ist die einfache Bestimmung und Auswertung der erhaltenen Hülle, mit einer entsprechend akzeptablen Rechenzeit. Von Nachteil ist jedoch, dass der Algorithmus nur eine konvexe Hülle bestimmt. Speziell bei Ottomotoren erweist sich die Darstellung als konvexe Hülle als unzureichend [85], da hier verstärkt nicht zulässige Stellgrößenkombinationen entstehen können. Diese mussten bisher manuell nachträglich angepasst werden.

Erweiterung auf konkave Hüllen

Der Quickhull-Algorithmus sowie der Delaunay-Triangulierung sind nur auf konvexe Räume beschränkt. Durch eine Erweiterung auf hochdimensionale konkave Räume können konkave Stellbereiche mathematisch beschrieben und automatisch im Rahmen einer Online-Vermessung berücksichtigt werden, ohne dass der Stellbereich mit Expertenwissen manuell angepasst werden muss.

Das Vorgehen zur Bestimmung einer konvexen Triangulation und Hülle ist Bild 3.12 a) dargestellt. Die Variationsraumvermessung erfolgt in Richtung beliebiger Stellgrößenkombinationen mithilfe eines Hüllrasters. Dabei wird zwischen aufeinanderfolgenden Messplanpunkten des

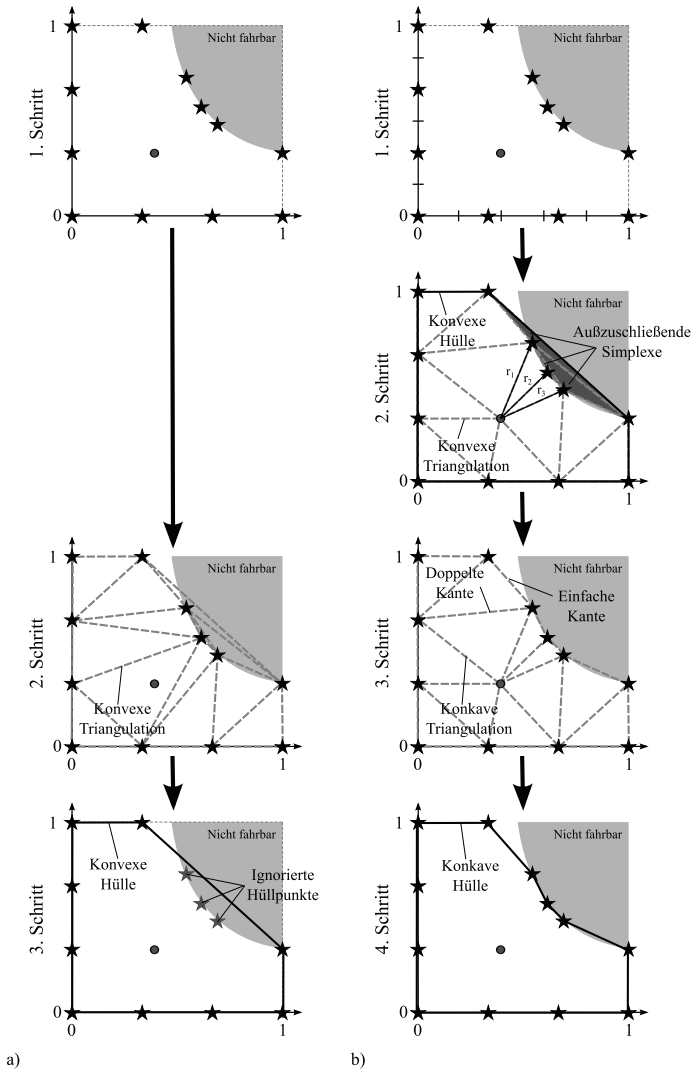


Bild 3.12: Vorgehen zur Bestimmung der die Punktemenge einschließenden a) konvexen und b) konkaven Hülle sowie Triangulation.

Hüllrasters jeweils zuerst der sichere Zentralpunkt angefahren (vgl. Bild 3.9 a)). Dadurch erfolgt die Vermessung des Variationsraums immer radial von einem sicheren Zentralpunkt aus. Ausgehend von der sich aus der Variationsraumvermessung ergebenden maximal zulässigen Stellgrößenkombinationen (1.Schritt) wird mithilfe der Delaunay-Triangulation ein konvexes

Dreiecksnetz erstellt (2. Schritt). Anschließend kann aus den Hüllpunkten mit dem Quickhull-Algorithmus die konvexe Hülle erstellt werden (3. Schritt). Diese enthält nicht fahrbare Punkte.

Die automatische Bestimmung einer konkaven Hülle, welche ausschließlich fahrbare Punkte beinhaltet, ist in Bild 3.12 b) dargestellt. Ausgehend von der Variationsraumvermessung (1. Schritt) wird dabei zunächst eine konvexe Delaunay-Triangulation mit einer Punktemenge bestehend aus allen Hüllpunkten sowie dem sicheren Zentralpunkt durchgeführt (2. Schritt). Die Bestimmung der konkaven Delaunay-Triangulation (3. Schritt) erfolgt durch Entfernen der fälschlicherweise zur konvexen Triangulation hinzugenommenen Dreiecke. Dies erfolgt mithilfe der konvexen Hülle. Der Quickhull-Algorithmus übergeht bei der Bestimmung der Hülle gewisse Punkte. Diese liegen zwar innerhalb der konvexen Hülle, sind aber im Rahmen der Variationsraumvermessung als hüllbildende Punkte bestimmt worden und somit relevant für die Bildung der konkaven Hülle. Diese somit fälschlicherweise bei der Hüllbildung ignorierten, innen-liegenden O Stellgrößenkombinationen $\mathbf{u}_o = (u_{1,o}, u_{2,o})$ mit $o = 1, \dots, O$ werden in Kugelkoordinaten transformiert:

$$\begin{bmatrix} u_{1,1} & u_{2,1} \\ \vdots & \vdots \\ u_{1,O} & u_{2,O} \end{bmatrix} \Rightarrow \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_O \end{bmatrix} \star \begin{bmatrix} r_1 \\ \vdots \\ r_O \end{bmatrix} \quad (3.18)$$

Durch die Transformation wird jede zweidimensionale Stellgrößenkombination \mathbf{u}_o , bestehend aus zwei kartesischen Koordinaten, stattdessen durch einen Radius r_o und einen Winkel φ_o beschrieben. Unter der Bedingung, dass jede Stellgrößenkombination ausgehend vom Zentralpunkt radial vermessen worden ist, kann der Bereich entlang der Strecke zwischen dem Zentralpunkt sowie der Stellgrößenkombination \mathbf{u}_o als innerhalb der Hülle liegend definiert werden. Die Lage der Strecke im Raum ergibt sich dabei aus dem Radius r_o kombiniert mit dem Winkel φ_o . Der über den Radius r_o hinausgehende Bereich gehört dagegen nicht zum zulässigen Stellbereich und liegt außerhalb der konkaven Hülle (siehe Schritt 2). Dieser auszuschließende Bereich kann bestimmt werden, indem der Radius r_o um einen Faktor $l > 0$ gestreckt wird, bei gleichbleibendem Winkel φ_o :

$$\begin{bmatrix} r_1^* \\ \vdots \\ r_O^* \end{bmatrix} = f \cdot \begin{bmatrix} r_1 \\ \vdots \\ r_O \end{bmatrix} \quad (3.19)$$

Diese so bestimmte radial verschobene Stellgrößenkombination \mathbf{u}_o^* liegt somit innerhalb eines Dreieckes der konvexen Triangulation, welches für die Bildung der konkaven Triangulation nicht zulässig ist. Diese Dreiecke wurden fälschlicherweise bei der Bestimmung der Delaunay-Triangulation hinzu genommenen und müssen somit ignoriert werden. Nach Reduktion der konvexen Delaunay-Triangulation um die so bestimmten, nicht zulässigen Dreiecke, ergibt sich die konkave Delaunay-Triangulation (3.Schritt).

Ausgehend von dieser kann die konkave Hülle bestimmt werden. Hierfür wird die Eigenschaft der Triangulation verwendet: Der Raum wird in mehrere zweidimensionale Dreiecke zerlegt, wobei jedes davon durch 3-Kanten definiert ist. Der sich aus allen Kanten der Dreiecke ergebende Datensatz beinhaltet somit doppelt vorkommende Kanten, welche im Inneren des Raum-

es liegen, sowie einmalig vorkommende Kanten, welche am Rand liegen (siehe 3. Schritt). Alle im gesamten Datensatz nur einmalig vorkommenden Kanten im gesamten Datensatz definieren somit die tatsächliche äußere konkave Hülle (4. Schritt).

Mithilfe der beschriebenen Vorgehensweise können konkave Räume dargestellt und in der integrierten Gesamtmethodik berücksichtigt werden, ohne dass ein manuelles Anpassen notwendig ist. Die nötigen Erweiterungen zum Einsatz in höheren Dimensionen sind in Anhang C zusammengefasst.

3.5.2 Dimensionsreduktion von Variationsräumen

Die in diesem Kapitel dargestellten Beispiele stellen zweidimensionale Variationsräume dar. Eine Erweiterung um weitere Dimensionen ist generell möglich (siehe Anhang C), jedoch wird die Bestimmung und Beschreibung der sich auf die Weise ergebenden Variationsräume immer rechenaufwendiger sowie immer konkaver.

Die Dimensionalität P eines globalen Variationsraumes kann reduziert werden, indem H Stellgrößen nicht in die Variationsraumbeschreibung integriert werden. Stattdessen werden L^H lokale Variationsräume der Dimension $P - H$ erstellt, welche auf L diskrete Stufen der herausgenommenen H Stellgrößen verteilt sind, siehe Bild 3.13. Die lineare Interpolation der umliegenden $P - H$ -dimensionalen Räume für Werte der H Stellgrößen zwischen den diskreten Stufen ermöglicht die Beschreibung des stellbaren Raumes für alle P Stellgrößen.

In Bild 3.13 ist der Zusammenhang zwischen globalen sowie lokalen und global-lokalen Variationsräumen dargestellt. Der globale Variationsraum mit $P = 3$ Dimensionen wird dabei um eine Stellgröße (u_3) reduziert. Die reduzierten lokalen, 2-dimensionalen Variationsräume werden im Anschluss auf L Stufen im Bereich $u_{3,\min} \dots u_{3,\max}$ verteilt.

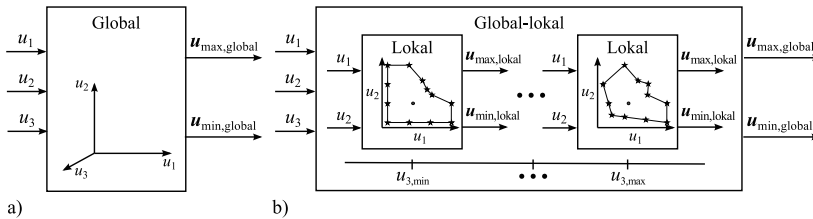


Bild 3.13: Darstellung eines 3-dimensionalen Variationsraums als a) globaler Variationsraum sowie b) 2-dimensionalen global-lokalen Variationsräumen. Bei den global-lokalen Variationsräumen ist die Stellgröße u_3 nicht in die Variationsraumbeschreibung integriert, sondern dient als Bezugsgröße zur Interpolation zwischen den Räumen.

Sowohl mit dem globalen als auch mit dem global-lokalen Variationsräumen können die maximal zulässigen Stellbereiche für alle beliebigen Stellgrößenkombinationen dargestellt werden.

Diese global-lokale Beschreibung des Variationsraums ermöglicht eine Reduktion der Dimensionalität der Räume, welche durch die Erhöhung der Anzahl an Räumen erreicht wird. Zu berücksichtigen ist dabei, dass eine lineare Reduktion der Dimensionalität eine exponentielle Erhöhung der Anzahl an Variationsräumen zur Folge hat. Weiterhin benötigen stark nichtlineare Zusammenhänge eine feinere Rasterung der reduzierten Räume. Demzufolge sollte generell eine möglichst hohe Dimensionalität der lokalen Variationsräume angestrebt werden.

3.5.3 Interpolation von Variationsräumen

Um überdies zulässige Stellgrößenkombinationen zwischen den diskreten Stufen bestimmen zu können, muss geeignet interpoliert werden. Durch eine lineare Interpolation der Hüllen aller umliegenden Variationsräume kann die maximal zulässige Stellgrößenkombination für jede beliebige Zwischenstufe bestimmt werden.

Jede Hülle eines Variationsraumes wird durch ihre Stützstellen, die maximal zulässigen Stellgrößenkombinationen \mathbf{u}_o , definiert. Wie in Bild 3.13 b) dargestellt, können die umliegenden Räume dabei aus einer unterschiedlichen Anzahl an Stützstellen bei unterschiedlichen Positionen bestehen.

Um eine Interpolation trotzdem eindeutig durchführen zu können, muss jede Stützstelle eines Variationsraumes in allen $j = 1 \dots J$ umliegenden lokalen Räumen Stützstellen in der gleichen Richtung haben. Die gleiche Richtungen entspricht dabei gleichen Winkeln φ_o bei der Kugelkoordinatendarstellung. Für jede Stützstelle muss in allen umliegenden Hüllen der sich ergebende Radius $r_{VRj,o}$ bei gleicher Stützstellenrichtung φ_o bestimmt werden.

Das Vorgehen ist exemplarisch in Bild 3.14 für den dreidimensionalen Fall mit u_3 als globale Stellgröße und u_1 sowie u_2 als lokale Stellgrößen dargestellt. Im ersten Schritt werden dabei die kartesischen Koordinaten der Stützstellen in ihre Kugelkoordinaten transformiert. Im Anschluss werden die in den benachbarten Variationsräumen VRj (mit $j = 1..2$ bei zwei benachbarten Räumen) gegenseitig fehlenden Stützstellen ergänzt. Dabei muss der Schnittpunkt zwischen dem Variationsraum und der sich sowohl aus dem Zentralpunkt als auch aus der Stützstelle ergebenden Geraden mittels zweier Geradengleichungen berechnet werden. Wesentlich ist, dass nicht die gesamte Delaunay-Triangulation auf einen möglichen Schnittpunkt überprüft werden muss, sondern nur die sich für den Variationsraum ergebende Hülle. Dadurch kann der Rechenaufwand wesentlich reduziert werden. Gemäß dem Beispiel aus Bild 3.14 entspricht jeder beliebige Punkt $s_{VRj,i}$ auf der Strecke zwischen dem Zentralpunkt z_{VRj} und dem Maximalpunkt $m_{VRj,o}$, bei Darstellung in Kugelkoordinaten, dem durch Verlängerung des Radius $r_{VRj,o}$ bis zur Grenze des Stellbereichs entstehenden Punktes. Folgende Gleichung gilt hierbei in Parameterform:

$$s_{VRj,o} = z_{VRj} + a \cdot (m_{VRj,o} - z_{VRj}). \quad (3.20)$$

Zusätzlich kann für jede Teilstrecke der Hülle, aufgespannt durch die zwei Stützstellen $\mathbf{u}_{VRj,o}$ und $\mathbf{u}_{VRj,o+1}$, die Geradengleichung in Parameterform aufgestellt werden:

$$s_{VRj,o} = \mathbf{u}_{VRj,o} + b \cdot (\mathbf{u}_{VRj,o+1} - \mathbf{u}_{VRj,o}) \quad (3.21)$$

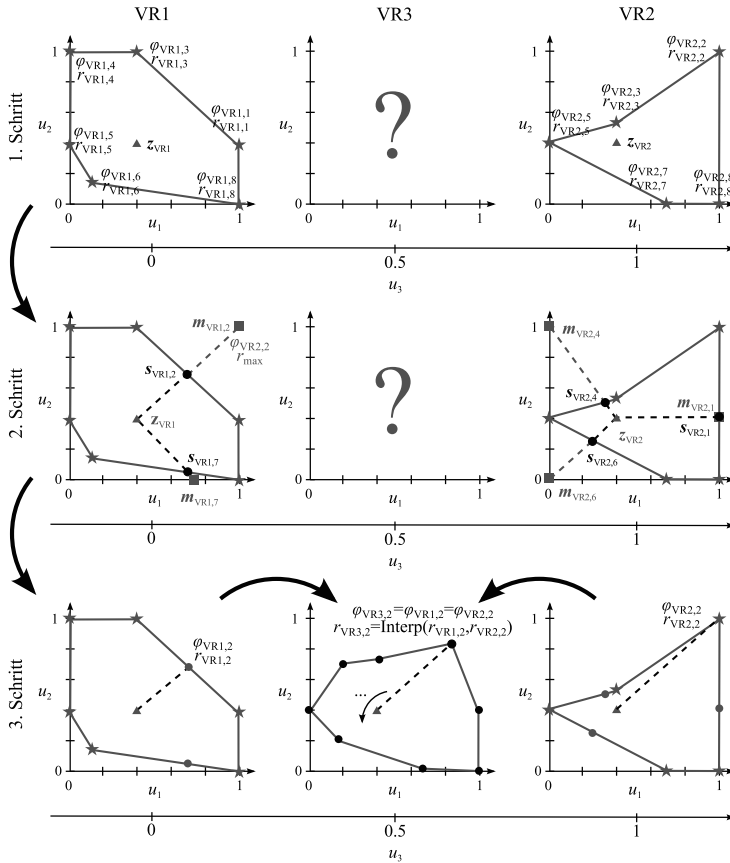


Bild 3.14: Vorgehen zur Bestimmung eines interpolierten Variationsraumes bzw. einer interpolierten Hülle. Nach der Transformation aller kartesischen Hüllpunkte in Kugelkoordinaten (1. Schritt), werden die gegenseitig fehlenden Hüllpunkte der Räume VR1 und VR2 bei gleichen Richtungen (2. Schritt) bestimmt. Abschließend können dann die interpolierten Hüllpunkte des gesuchten Variationsraums VR3 durch Interpolation ermittelt werden (3. Schritt).

Durch Gleichsetzen und Umstellen folgt

$$\mathbf{z}_{VRj} - \mathbf{u}_{VRj,o} = a \cdot (\mathbf{z}_{VRj} - \mathbf{m}_{VRj,o}) + b \cdot (\mathbf{u}_{VRj,o+1} - \mathbf{u}_{VRj,o}) \quad (3.22)$$

bzw. in Matrixschreibweise

$$[\mathbf{z}_{VRj} - \mathbf{u}_{VRj,o}] = [\mathbf{z}_{VRj} - \mathbf{m}_{VRj,o}, \mathbf{u}_{VRj,o+1} - \mathbf{u}_{VRj,o}] \cdot \begin{bmatrix} a \\ b \end{bmatrix}. \quad (3.23)$$

Durch Bestimmung von a und b mithilfe der Matrixinversion

$$\begin{bmatrix} a \\ b \end{bmatrix} = [\mathbf{z}_{VRj} - \mathbf{m}_{VRj,o} \cdot \mathbf{u}_{VRj,o+1} - \mathbf{u}_{VRj,o}]^{-1} \cdot [\mathbf{z}_{VRj} - \mathbf{u}_{VRj,o}] \quad (3.24)$$

kann der Schnittpunkt der Strecke $\overline{\mathbf{z}_{VRj} \mathbf{m}_{VRj,o}}$ mit jeder durch $\overline{\mathbf{u}_{VRj,o} \mathbf{u}_{VRj,o+1}}$ aufgespannten Teilstrecke der Hülle bestimmt werden. Die Lösung, für die $a, b \in [0, 1]$ gilt, definiert den gesuchten Schnittpunkt und somit die fehlende Stützstelle in dem Variationsraum.

Nachdem die Räume VRj mithilfe des beschriebenen Vorgehens gegenseitig um die fehlenden Stützstellen der gleichen Richtung ergänzt worden sind, kann aus den umliegenden Stützstellen gleicher Richtung die lineare Interpolation erfolgen (3. Schritt in Bild 3.14). Da sich die Stützstellen ausschließlich in den Radien $r_{VRj,o}$, jedoch nicht in den Winkeln φ_o unterscheiden, muss für den gesuchten Zwischenraum nur der Radius linear interpoliert werden. Bei kartesischen Koordinaten müssten an dieser Stelle alle Dimensionen einzeln interpoliert werden, was den doppelten Rechenaufwand zur Folge hätte. Der aus den zwei umliegenden Variationsräumen $VR1$ sowie $VR2$ interpolierte Radius $r_{VR3,i}$ des Variationsraums $VR3$ berechnet sich gemäß der linearen Interpolation zu:

$$r_{VR3,o}(u_{3,VR3}) = r_{VR1,o} \frac{u_{3,VR2} - u_{3,VR3}}{u_{3,VR2} - u_{3,VR1}} + r_{VR2,o} \frac{u_{3,VR3} - u_{3,VR1}}{u_{3,VR2} - u_{3,VR1}}. \quad (3.25)$$

Entsprechend können alle interpolierten Radien $r_{VR3,o}$ für jede gesuchte Stützstelle $o = 1, \dots, O$ ermittelt werden. Zusammen mit den zugehörigen Winkeln definiert der interpolierte Radius den maximal möglichen Stellbereich bzw. Variationsraum in dem gesuchten Zwischenraum $VR3$. Eine Erweiterung dieses Vorgehens auf höhere Dimensionen in Kombination mit zwei globalen Stellgrößen ist durch Anwenden von entsprechend dimensionalen Geraden- und Ebenengleichungen sowie der bilinearen Interpolation möglich. Das sich dafür ergebende Vorgehen ist in Anhang C zusammengefasst beschrieben.

Liegen die global-lokalen Hüllen in der globalen Dimension weit auseinander, kann der Fall auftreten, dass die Hüllen zueinander verschoben sind. Dabei ist keine gemeinsame lokale Stellgrößenkombination vorhanden, welche in allen Räumen stellbar ist. Dadurch werden die Räume ausgehend von verschiedenen zentralen Startpunkten vermessen. Bei der anschließenden Interpolation muss in diesem Fall zusätzlich der Zentralpunkt \mathbf{z}_{VR3} für den zu bestimmenden Raum $VR3$ aus den beiden umliegenden Zentralpunkten interpoliert werden. Die Interpolation der Radien erfolgt dabei weiterhin nach Formel (3.25).

In Bild 3.15 sind exemplarisch die vermessenen Hüllen für drei Motorarbeitspunkte ($2000 \text{ min}^{-1} / 15 \frac{\text{mm}^3}{\text{Hub}}$), ($2000 \text{ min}^{-1} / 20 \frac{\text{mm}^3}{\text{Hub}}$) und ($2000 \text{ min}^{-1} / 25 \frac{\text{mm}^3}{\text{Hub}}$) dargestellt. Als Stellgrößen wurden u_{HDAGR} sowie u_{VTG} sowie als Grenzen $100 \text{ ppm} \leq \text{NO}_x \leq 700 \text{ ppm}$ und kein Auftreten von Zündaussetzern verwendet. Weiterhin ist für die mittlere Einspritzmenge von $20 \frac{\text{mm}^3}{\text{Hub}}$ der aus den umliegenden Räumen interpolierte Variationsraum dargestellt. Die Fläche des mittleren Arbeitspunktes für den vermessenen Raum beträgt 41.19 % und die Fläche des interpolierten Raumes beträgt 39.74 %. Das Vermessen des dritten Variationsraumes bringt in diesem Fall eine Verbesserung von 1.45 %, wobei ein zusätzlicher Messaufwand von 479 s aufgewendet werden muss. Mithilfe der Interpolation kann somit ca. 1/3 der Messzeit eingespart werden, bei einer ausreichend genauen Beschreibung des Variationsraums bei ($2000 \text{ min}^{-1} / 20 \frac{\text{mm}^3}{\text{Hub}}$).

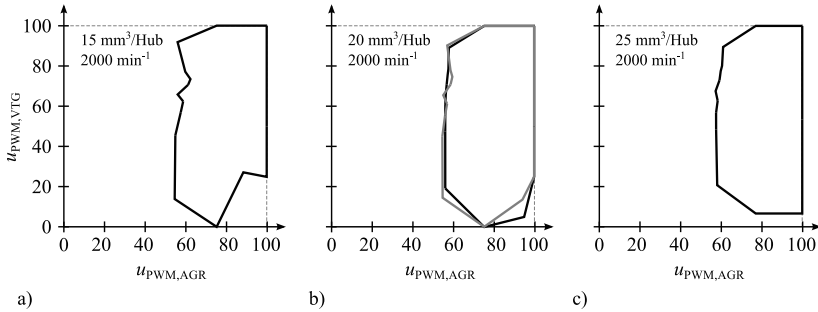


Bild 3.15: Vermessene Hüllen für die Arbeitspunkte ($2000 \text{ min}^{-1} / 15 \frac{\text{mm}^3}{\text{Hub}}$), ($2000 \text{ min}^{-1} / 20 \frac{\text{mm}^3}{\text{Hub}}$) und ($2000 \text{ min}^{-1} / 25 \frac{\text{mm}^3}{\text{Hub}}$). Für die Stellgrößen u_{HDAGR} sowie u_{VTG} ergeben sich die dargestellten Räume aufgrund der Grenzen $100 \text{ ppm} \leq \text{NO}_x \leq 700 \text{ ppm}$ und keine Zündaussetzer. Zusätzlich ist der berechnete Zwischenraum ($2000 \text{ min}^{-1} / 20 \frac{\text{mm}^3}{\text{Hub}}$) dargestellt, der eine gute Übereinstimmung mit dem vermessenen Raum aufweist.

3.6 Zusammenfassung

Die in diesem Kapitel beschriebene Vorgehensweise zur Variationsraumvermessung ermöglicht eine schnellere Bestimmung und einfachere Beschreibung von konkaven Stellbereichen als bei klassischen Offline-Verfahren. Die sich ergebenden Variationsräume berücksichtigen dabei weiche Grenzgrößen, die entweder direkt gemessen (z.B. Emissionen) oder indirekt ermittelt (z.B. Zündaussetzer) werden.

Die Berücksichtigung von Zündaussetzern wird mit einem signalbasierten Verfahren realisiert, welches die echtzeitfähige Zylinderdruckindizierung nutzt. Die statistische Auswertung der maximalen Zylinderdruckdifferenzen ermöglicht die Definition einer vom Arbeitspunkt unabhängigen Erkennungsschwelle.

Die echtzeitfähige Vermessung und Berücksichtigung aller relevanten Größen ermöglicht die Integration der Variationsraumvermessung in eine automatisierte Vermessung am Prüfstand. Die Größen können dabei diskret oder kontinuierlich bzw. direkt messbar oder indirekt bestimmbar sein. Durch dieses Vorgehen erfolgt die Vermessung ohne Eingriffe eines Prüfstands-fahrers.

Die Vermessung des Variationsraums erfolgt mit einem Hüllraster als Messplan, wodurch $(L - 2)^P$ -Messpunkte gegenüber einem klassischen Raster eingespart werden. Der sich aus dem Hüllraster ergebende zulässige Variationsraum wird durch eine Grenzraumregelung ermittelt. Diese legt die Aktorstellgeschwindigkeit fest, basierend auf dem Abstand aller Grenzgrößen zu Ihren zulässigen Limits. Die Dauer der Vermessung ist dabei stark von der Anzahl an zu bestimmenden Hüllrasterpunkten sowie den zugrundeliegenden physikalischen Prozessen abhängig und kann mit bis zu ca. 50 % Zeitersparnis erfolgen.

Basierend auf den vermessenen Grenzpunkten wurde eine einfache konkave mathematische Beschreibung des zulässigen Stellbereichs entwickelt. Letztere baut auf der Delaunay-Triangulation und dem Quickhull-Algorithmus auf. Diese werden klassischerweise zur Bestimmung von konvexen Räumen genutzt, und es existieren hierfür schnelle Algorithmen.

Die Komplexität von Variationsräumen bei höherdimensionalen Vermessungen wird durch eine Dimensionsreduktion sowie Interpolation der Räume erreicht. Dabei wird die Anzahl an Stellgrößen reduziert, indem aus dem globalen Variationsraum mehrere global-lokale Räume abgeleitet werden. Durch die Kombination mit einer Interpolation von konkaven Zwischenräumen kann die Anzahl der entstehenden Variationsräume weiter reduziert werden.

4 Online-Versuchsplanung bei der Motorvermessung

Eine geeignete Versuchsplanung muss auf den zugrundeliegenden Prozess ausgerichtet sein, was bislang ein entsprechendes Wissen über diesen voraussetzt. Durch den Einsatz von aktiv lernenden Versuchsplänen können auch unbekannte Prozesse zielgerichtet vermessen werden. Für die Erstellung eines Messplans wird ein kombiniertes Verfahren entwickelt, welches in einem iterativ angepassten Stellbereich sowohl die Raumfüllung als auch Modelleigenschaften bewertet. Dabei werden neue Messpunkte automatisiert an die Stellen platziert, an denen der Informationsgewinn für die folgende Modellbildung und/oder Stellgrößenoptimierung am größten ist. Bei einer dynamischen Versuchsplanung spielen zusätzlich die Wahl der Abtastzeit sowie eines geeigneten Anregungssignals eine wesentliche Rolle. Durch die Wahl einer signalbasierten Abtastzeit kann das Maximum an Information aus den Signalen automatisiert extrahiert werden. Kombiniert mit einem prozessangepassten deterministischen Rechtecksignal, welches tiefe Frequenzen stark anregt, kann der Prozess flexibel angeregt werden. Dadurch kann bereits mit einer kurzen Signallänge eine hohe statische sowie dynamische Güte des identifizierten Modells erreicht werden.

Bei der Vermessung von Motormodellen kommen Methoden der statistischen und modellbasierten Versuchsplanung zum Einsatz. Dabei wird mithilfe von statistischen Verfahren die Durchführung von wissenschaftlichen Experimenten bzw. Versuchen effizienter gestaltet [38]. Je nach zu vermessender Ausgangsgröße ist die Anzahl an benötigten Messpunkten abhängig vom Modelltyp und der gewünschten Anzahl an Wiederholmessungen. Die Schwierigkeit besteht darin, die Anzahl an geeigneten Messpunkten im Voraus festzulegen, da zu diesem Zeitpunkt weder Modelltyp noch Störgrößen etc. bekannt sind. Weiterhin sind bei der Vermessung an Motorenprüfständen nicht nur das stationäre Prozessverhalten, sondern auch die dynamischen Zusammenhänge von Interesse. Um diese ausreichend für die Identifikation abzubilden, müssen geeignete Anregungssignale gewählt werden. Durch eine zielgerichtete Versuchsplanung kann die Anzahl an benötigten Messpunkten und somit die Dauer der Vermessung stark beeinflusst werden. Im Rahmen der Online-Vermessung werden iterative Methoden der statistischen Versuchsplanung eingesetzt, um die möglichst kleinste Anzahl an Messpunkten zu erhalten und somit die kürzeste Messdauer zu erzielen.

Im Folgenden werden die Details des zweiten Schritts im Rahmen einer Vermessung (siehe Bild 4.1) vorgestellt. Nach einem kurzen Überblick über die Grundlagen der statistischen Versuchsplanung und Optimierung (siehe Abschnitt 4.1) wird ein neues Anregungssignal vorgestellt, mit welchem sowohl statische als auch dynamische Modelle bestimmt werden können (siehe Abschnitt 4.2). Anschließend folgt eine Beschreibung der Methodik zur Erstellung

von zielgerichteten Messplänen mithilfe einer Stellbereichsadaption sowie modellbasierten und messdatenbasierten Auswertung (siehe Abschnitt 4.3).

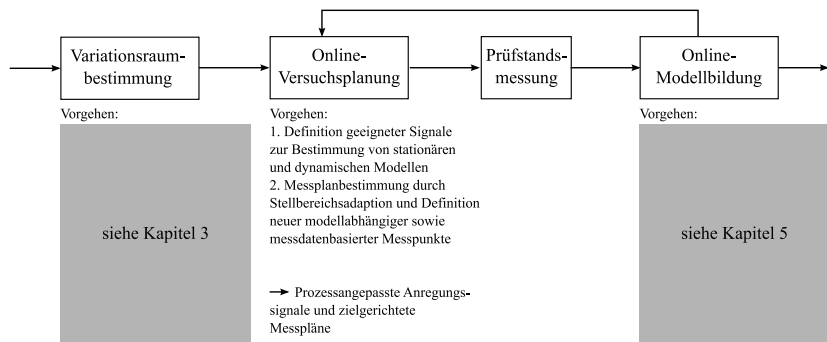


Bild 4.1: Vereinfachte Darstellung der einzelnen Schritte einer Vermessung und daraus resultierend das Vorgehen zum Entwurf einer durchgängigen Online-Methode im Rahmen dieser Arbeit bzw. in diesem Kapitel.

4.1 Grundlagen der statistischen Versuchsplanung und modellbasierten Optimierung

Angelehnt an [66, 110] umfasst die statistische Versuchsplanung (Design of Experiments, DoE-Methoden) die Einflussgrößenauswahl (Bestimmung signifikanter Prozessstellgrößen auf den Prozessausgang), die Effektermittlung (Bestimmung der Einflüsse der Stellgrößen untereinander) und die Messpunktplatzierung im Eingangsraum. Im Rahmen einer dynamischen Vermessung, muss die statistische Versuchsplanung zusätzlich um die Auswahl einer geeigneten Prozessanregung sowie Abtastzeit ergänzt werden. Sind die physikalischen Zusammenhänge des zu modellierenden Prozesses bekannt, können daraus die relevanten Einflussgrößen und Effekte direkt abgeleitet werden. In diesem Fall reduziert sich die statistische Versuchsplanung auf die gezielte Messpunktverteilung im Eingangsraum sowie die Auswahl eines adäquaten dynamischen Anregungssignals und der entsprechenden Abtastzeit. Ziel dabei ist, den Informationsgehalt des geplanten Experiments bei einer minimalen Anzahl an Messpunkten zu maximieren. Anstelle eines vollständigen Versuchsplanes wie bei einer Rastervermessung muss in diesem Fall nur noch ein reduzierter Messplan aufgestellt werden, welcher die zu vermessenden Messpunkte enthält.

4.1.1 Arten der Datenaufzeichnung

Bei der Vermessung von Motormodellen wird zwischen verschiedenen Arten der Messdatenaufzeichnung unterschieden. Je nach Einsatzgebiet des zu bestimmenden Modells und der Prüf-

standsausstattung (Aktorik, Messtechnik und Automatisierungssystem) wird zwischen einer stationären oder einer dynamischen Vermessung unterschieden.

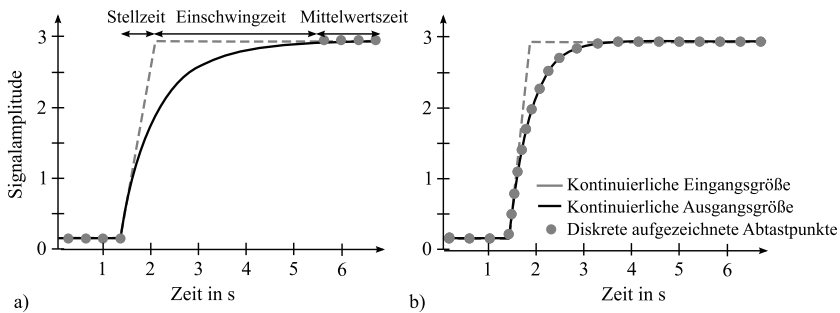


Bild 4.2: Vergleich einer statischen und einer dynamischen Datenaufzeichnung. a) Statische Datenerfassung mit Aufzeichnung des Mittelwerts im eingeschwungenen Zustand sowie b) dynamische Aufzeichnung mit einer festen Abtast-Schrittweite

Stationäre Datenaufzeichnung

Bei einer stationären Datenaufzeichnung werden alle Stellgrößen durch Sollwertvorgabe auf neue diskrete Zielgrößen verstellt und der eingeschwungene Zustand aller Ausgangs- und Zustandsgrößen abgewartet. Das Verstellen kann dabei rampen- oder sprungförmig erfolgen, wobei letzteres schneller erfolgt jedoch das Risiko eines Überschwingens mitbringt. Im eingeschwungenen Zustand aller Größen wird dann für eine zu definierende Zeitdauer die Aufzeichnung der abgetasteten Messgrößen aktiviert und aus diesen ein (zeitlicher) Mittelwert gebildet. Dadurch können sowohl Zyklusschwankungen als auch Störsignale des Verbrennungsmotors eliminiert bzw. reduziert werden [19]. Mit dem sich aus aufgezeichneten Messdaten definierten stationären Modell kann das stationäre Prozessverhalten des Motors abgebildet werden. Die Vermessung von stationären Modellen kann ausschließlich auf Motor- oder Rollenprüfständen durchgeführt werden, da bei diesen Umgebungen Versuchsbedingungen und Eingangsgrößen konstant gehalten werden können. Das Haupteinsatzgebiet der stationären Motorvermessung ist die Bestimmung von stationären optimierten Sollwertkennfeldern.

Dynamische Datenaufzeichnung

Eine weitere Möglichkeit besteht in der dynamischen Datenaufzeichnung. Im Gegensatz zur stationären Datenerfassung erfolgt dabei eine durchgehende Datenaufzeichnung, indem der Datenrekorder mit einer festen Abtast-Schrittweite getaktet wird. Dadurch kann zusätzlich das Prozessverhalten in den Übergängen von einer Stellgrößenkombination in eine andere erfasst werden.

Mit dem aus den dynamisch aufgezeichneten Messdaten ermittelten dynamischen Modell kann das dynamische Prozessverhalten bestimmt werden. Dieses wird vor allem durch die Dynamik einzelner Motorkomponenten (Aktoren, Speichervorgänge) sowie die Dynamik der Medienkreisläufe (Öl-, Kühlmittel und Kraftstoff) bestimmt. Die Vermessung eines dynamischen Modells kann sowohl auf Motor- oder Rollenprüfständen als auch in realen Fahrzeugen erfolgen. Bei der Vermessung in realen Fahrzeugen müssen die nicht konstanten Versuchsbedingungen bei der Modellierung einbezogen werden. Die dynamische Vermessung kommt bei dem Entwurf und der Applikation von Funktionen zum Einsatz, welche für den optimalen Motorbetrieb in dynamischen Fahrzyklen oder Betriebsarten zuständig sind. Hierfür werden dynamische (Teil)-Modelle genutzt, um das dynamische Prozessverhalten zu simulieren.

Insbesondere dynamische Messdaten sind in der Regel verrauscht, Störungen unterworfen und werden hochaufgelöst aufgezeichnet, was zu einer hohen Datenmenge führt. Sehr kleine Abtastzeiten führen dabei zu numerischen Problemen [70]. Die Wahl der idealen Abtastzeit für die Filterung bei der dynamischen Modellbildung kann mithilfe der folgenden Faustformel, unter Annahme eines linearen Prozesses mit Tiefpassverhalten, bestimmt werden

$$T_0 \approx T_{95} \frac{1}{5} \dots T_{95} \frac{1}{15} \quad (4.1)$$

wobei T_0 die Abtastzeit und T_{95} die Zeit ist, bei der 95% vom Endwert erreicht ist [70]. Liegt dagegen ein schwingungsfähiger Prozess oder ein Prozess mit mehreren verschiedenen Zeitkonstanten vor, muss die Abtastzeit nach [173] noch kleiner gewählt werden. Dabei wird eine Parametrierung in Abhängigkeit von der kleinsten Prozesszeitkonstante T_{\min} vorgeschlagen:

$$T_0 \approx T_{\min} \frac{1}{5} \quad (4.2)$$

Die Faustformeln haben den Nachteil, auf Tiefpassprozesse mit bekannten Zeitkonstanten beschränkt zu sein. Im Rahmen der Online-Schätzung ist jedoch der zugrundeliegende Prozess sowie dessen Zeitkonstante unbekannt, wodurch die passende Parametrierung der Abtastzeit vor Beginn der Vermessung nicht möglich ist. Aus diesem Grund wurde im Rahmen der Online-Methodik ein heuristischer Ansatz umgesetzt, welcher in Anhang E beschrieben ist.

4.1.2 Einflussgrößenauswahl und Effektermittlung

Einflussgrößen werden bei der statistischen Versuchsplanung in konstante (z.B. konditionierte Kraftstofftemperatur), zufällige sowie einstellbare Größen unterteilt. Zufällige Größen sind dabei immer quantitativ (z.B. Umgebungsdruck und -temperatur), während einstellbare Einflussgrößen qualitativ (z.B. verschiedene Kraftstoffarten) oder quantitativ (z.B. unterschiedliche Haupteinspritzmengen) sein können. Die Einflussgrößen können dabei direkt oder in Wechselwirkung mit weiteren Größen den Prozessausgang beeinflussen. Bei dem klassischen DoE müssen diese Annahmen vorneweg in Form eines Modelles getroffen werden, um im Anschluss mit der Regressionsanalyse die Modellparameter zu schätzen und die Annahmen zu überprüfen.

Im Rahmen der hier vorgestellten Online-Methodik wird ein alternatives Vorgehen angewandt. Ausgehend von den Messdaten der Variationsraumvermessung wird ein erstes Modell mit automatischer Regressorauswahl (Effektermittlung) identifiziert. Demzufolge existiert bereits vor der ersten statistischen Versuchsplanung ein stationäres Modell, dessen Einflussgrößen (Stellgrößen) vorgegeben und Effekte bestimmt worden sind. Die Messpunktverteilung kann dadurch direkt für das bereits identifizierte, grobe Modell erfolgen. Durch die iterative Vorgehensweise der Methodik werden die Effekte rekursiv an den Prozess angepasst, sodass weniger Annahmen nötig sind.

4.1.3 Messpunktverteilung

Die gezielte Messpunktverteilung im Stellgrößenraum und folglich die Auswahl der Amplituden aller Stellgrößen erfolgt mit klassischen, raumfüllenden oder modellbasierten Messplänen. Bei den meisten dieser Messpläne muss vorneweg ein Kandidatensatz K an möglichen Stellgrößenkombinationen erstellt werden, welcher den zulässigen Variationsraum gleichmäßig mit Kandidaten abdeckt. Ein *Kandidat* entspricht dabei einer zulässigen Stellgrößenkombination.

Die Erstellung des Kandidatensatzes K erfolgt entweder durch eine würfelförmige Rasterung oder durch eine sphärische Rasterung [138] aller Stellgrößen zwischen ihren minimalen und maximalen Werten. In Bild 4.3 sind die würfelförmige sowie sphärische Rasterung des Variationsraums aus Bild 3.9 dargestellt. Die würfelförmige Rasterung erfolgt mit 7 Teilungen je Stellgröße und die sphärische mit 12 Winkelteilungen sowie 5 Radianteilungen. Es wird ersichtlich, dass mit der würfelförmigen Rasterung 18 Kandidaten (28 %) unzulässig sind, während bei der sphärischen Rasterung nur 9 Kandidaten (15 %) nicht stellbar sind. Vorteile der sphärischen Rasterung sind, neben der geringeren Rechenzeit, die Erzeugung vieler geeigneter Kandidaten bei bereits grober Diskretisierung bzw. hoher Dimensionalität sowie die gute Eignung für langgezogene Variationsräume [138].

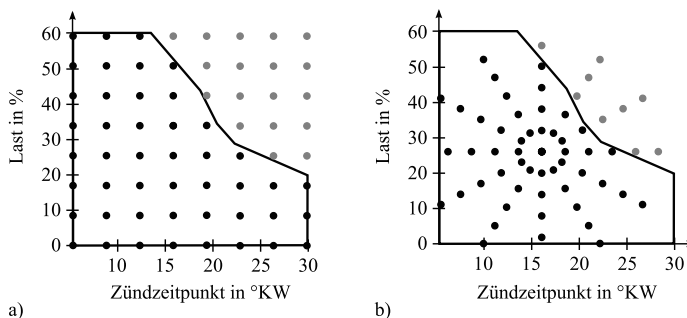


Bild 4.3: Vergleich einer würfelförmigen und sphärischen Rasterung für den Variationsraum aus Bild 3.9.

Aus dem erstellten Kandidatensatz K werden dann die tatsächlich zu stellenden Kandidaten M mit

$$M \in K \quad (4.3)$$

ausgewählt, welche ein dem Messplan entsprechendes Kriterium erfüllen.

Klassische Messpläne

Zu den klassischen Messplänen zählen die (würfelförmige oder sphärische) Rasterung sowie zentral zusammengesetzte Messpläne. Eine rasterförmige Abdeckung des Versuchsraumes zwischen den minimalen und maximalen Werten jeder Stellgröße ist aufgrund des Fluchs der Dimensionen nur bei kleiner Dimension P sowie wenigen Stufen L sinnvoll. Die Anzahl der Messpunkte N wird nach Gleichung (3.10) berechnet. Zentral zusammengesetzte Messpläne kombinieren einen würfelförmigen mit einem sternförmigen Messplan, indem die Mittelpunkte aufeinander gelegt werden. Nachteilig ist, dass hierbei ausschließlich der Zentralpunkt sowie Randpunkte vorgegeben werden. Diese Messpläne eignen sich somit nur für lineare Prozesse, da bei stark nichtlinearen Prozessen keine Prozessinformation in den Zwischenbereichen erfasst werden. Die klassischen Messpläne kommen ohne die Erstellung eines Kandidatensatzes aus, da in deren Fall gilt:

$$M = K \quad (4.4)$$

Raumfüllende Messpläne

Raumfüllende Messpläne decken den von den Stellgrößenkombinationen aufgespannten Raum möglichst gleichmäßig ab. Dadurch eignet sich ihr Einsatz im Falle von hoch-dimensionalen Versuchsräumen bei gleichzeitig unbekanntem Modellansatz. Raumfüllende Messpläne wählen aus dem gesamten Kandidatensatz an Stellgrößenkombinationen K zufällig einen Messplan M mit O Stellgrößenkombinationen.

Bei LHS-Messplänen (latin hypercube sampling, [104, 134]) handelt es sich um Zufallsstellgrößen, welche bei der Projektion auf die jeweiligen Stellgrößenachsen eine Gleichverteilung bei maximiertem Punktabstand oder minimaler Korrelation ergeben [89]. Dabei wird jede Dimension des Variationsraums in O äquidistante Bereiche unterteilt. Aus den daraus resultierenden O^P Hyperquadern wird iterativ zufällig eine Untermenge O ausgewählt, welche für jedes Aufteilungsintervall genau einen Hyperquader enthält. Innerhalb dieser Untermenge wird in jedem der O Hyperquader, welcher mehrere Kandidatenpunkte enthält, ein Kandidatenpunkt so ausgewählt, dass die Summe der O euklidischen Abstände zwischen den Kandidatenpunkten maximal wird. Dieses Vorgehen wird iterativ wiederholt, bis ein Abbruchkriterium erreicht wird (Summe aller euklidischer Abstände oder maximale Anzahl an Iterationen Z).

Weitere Methoden zur Generierung von raumfüllenden Messplänen sind Halton- [56], Faure- [37] und Sobol-Sequenzen [149]. Diese quasizufälligen Sequenzen sind Zahlenfolgen geringer Diskrepanz (gute Gleichverteilung), welche speziell für hoch-dimensionale Räume geeignet sind, da sie diese gleichmäßiger abdecken als Zufallszahlen [95, 99].

Modellbasierte Messpläne

Modellbasierte Messpläne erstellen zu vermessende Stellgrößenkombinationen basierend auf einem Modell, wie z.B. Polynomen. Annahmen sind dabei, dass sich der wahre Prozess durch das Modell darstellen lässt und die Messfehler unabhängig identisch normalverteilt sind (weißes Rauschen). Die sogenannten optimalen Messpläne stellen eine geschlossene Theorie zur Versuchsplanung zur Verfügung. *Optimal* bezieht sich dabei auf den Informationsgehalt eines Messplans, welcher mit möglichst wenigen Messpunkten maximiert wird [51]. Ausgehend von der Systemgleichung mit einem Störsignal \mathbf{n} sowie der Datenmatrix Ψ mit L Stellgrößen sowie N diskreten Stützstellenwerten (Stellgrößenkombinationen)

$$\mathbf{y} = \Psi \Theta + \mathbf{n} \quad (4.5)$$

$$\Psi = \begin{bmatrix} 1 & u_{1,1} & u_{2,1} & \cdots & u_{L,1} \\ 1 & u_{1,2} & u_{2,2} & \cdots & u_{L,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_{1,N} & u_{2,N} & \cdots & u_{L,N} \end{bmatrix} \quad (4.6)$$

werden nach [120] mit der Schätzung nach der Methode der kleinsten Quadrate die Modellparameter zu

$$\hat{\Theta} = \left(\Psi^T \text{cov} \{ \mathbf{n} \}^{-1} \Psi \right)^{-1} \Psi^T \text{cov} \{ \mathbf{n} \}^{-1} \mathbf{y} \quad (4.7)$$

bestimmt. Die Kovarianzmatrix der bestimmten Modellparameter

$$\text{cov} \{ \hat{\Theta} \} = \left(\Psi^T \Psi \right)^{-1} \Psi^T \text{E} \{ \mathbf{n} \mathbf{n}^T \} \Psi \left(\Psi^T \Psi \right)^{-1} \quad (4.8)$$

wird dazu genutzt, eine Aussage über die Genauigkeit der Schätzung zu treffen. Je kleiner die Kovarianzmatrix ist, desto kleiner ist der Schätzfehler der Modellparameter. Weiterhin gilt für die Varianz der modellierten Ausgangsgröße an einer Stelle $\mathbf{u}_i \in \Psi$:

$$\begin{aligned} \text{var} \{ \hat{\mathbf{y}} \} &= \text{diag} \left(\text{cov} \{ \hat{\mathbf{y}} \} \right) \\ &= \text{diag} \left(\mathbf{u}_i \text{cov} \{ \hat{\Theta} \} \mathbf{u}_i^T \right) \\ &= \text{diag} \left(\mathbf{u}_i \left(\Psi^T \Psi \right)^{-1} \Psi^T \text{E} \{ \mathbf{n} \mathbf{n}^T \} \Psi \left(\Psi^T \Psi \right)^{-1} \mathbf{u}_i^T \right). \end{aligned} \quad (4.9)$$

Durch Annahme eines weißen Rauschens als Störsignal \mathbf{n} mit Varianz σ^2 folgt für den Erwartungswert $\text{E} \{ \mathbf{n} \mathbf{n}^T \} = \sigma^2 \mathbf{I}$, wodurch sich die Gleichungen (4.8) und (4.9) zu

$$\text{cov} \{ \hat{\Theta} \} = \sigma^2 \left(\Psi^T \Psi \right)^{-1} \quad (4.10)$$

$$\text{var} \{ \hat{\mathbf{y}} \} = \text{diag} \left(\mathbf{u}_i \sigma^2 \left(\Psi^T \Psi \right)^{-1} \mathbf{u}_i^T \right) \quad (4.11)$$

vereinfachen. Hieraus wird ersichtlich, dass bei Übereinstimmung der Modell- mit der Prozessstruktur die Kovarianz der Modellparameter bzw. die Varianz des Modellausgangs ausschließlich von der Varianz des Messrauschens σ^2 sowie der Matrix $\left(\Psi^T \Psi \right)^{-1}$ abhängt. Das Messrauschen kann durch Verbesserung der Versuchsbedingung und -umgebung verbessert werden,

indem z.B. Störquellen beseitigt werden sowie bessere Messtechnik verwendet wird. Die Inverse der Fisher-Informationsmatrix $(\Psi^T \Psi)^{-1}$ beinhaltet dagegen die aus den Stellgrößen bestimmten Regressoren und kann durch Positionierung dieser im zulässigen Variationsraum beeinflusst werden. Für die Positionierung der Regressoren, und daraus folgend die Minimierung der (Kovarianz-/Varianz-)Matrix, können verschiedene Kriterien eingesetzt werden, aus denen sich direkt die Messpläne ableiten [7, 110, 123]:

- A-optimal: Minimierung der Spur

$$\rightarrow M = \arg \min_{\Psi \in K} \left(\text{trace} \left((\Psi^T \Psi)^{-1} \right) \right)$$

- C-optimal: Minimierung für einen vorzugebenden Vektor $c \in \mathbb{R}$

$$\rightarrow M = \arg \min_{\Psi \in K} \left(c^T (\Psi^T \Psi)^{-1} c \right)$$

- D-optimal: Minimierung der Determinanten

$$\rightarrow M = \arg \min_{\Psi \in K} \left(\det \left((\Psi^T \Psi)^{-1} \right) \right)$$

- E-optimal: Minimierung des größten Eigenwerts

$$\rightarrow M = \arg \min_{\Psi \in K} \left(\lambda_{\max} \left((\Psi^T \Psi)^{-1} \right) \right)$$

- G-optimal: Minimierung der maximalen Ausgangsgrößenvarianz für ein $u_i \in \Psi$

$$\rightarrow M = \arg \min_{\Psi \in K} \left(\arg \max \left(u_i^T (\Psi^T \Psi)^{-1} u_i \right) \right)$$

- I-optimal: Minimierung des Integrals über die Ausgangsgrößenvarianz für alle $u_i \in \Psi$ und die Dichte p auf Ψ

$$\rightarrow M = \arg \min_{\Psi \in K} \left(\int_{\Psi} u_i^T (\Psi^T \Psi)^{-1} u_i p(du_i) \right).$$

- V-optimal: Minimierung der durchschnittlichen Ausgangsgrößenvarianz für alle $u_i \in \Psi$

$$\rightarrow M = \arg \min_{\Psi \in K} \left(\frac{1}{N} \sum_{i=1}^N u_i^T (\Psi^T \Psi)^{-1} u_i \right)$$

Die Kriterien A-, C-, D- und E-optimal minimieren dabei die Kovarianz der Modellparameter, während die G-, I- sowie V-optimalen Kriterien die Varianz des Modellausgangs minimieren. Der D-optimale Messplan spannt dabei das größtmögliche Volumen über den Versuchsraum auf [32]. Im Gegensatz dazu wird bei einem A-optimalen Messplan die durchschnittliche Kovarianz der Modellparameter minimiert sowie beim G- und V-optimalen Messplan die maximale bzw. durchschnittliche Varianz des Modellausgangs minimiert.

In der Regel sind die Optimalitätskriterien nicht geschlossen lösbar. Die Erstellung von Messplänen unter Einbezug dieser Kriterien erfolgt durch Austauschalgorithmen [107]. Dabei werden, ausgehend von einem initialen Messplan M_0 , solange iterativ mit einem Kandidatensatz K einzelne Punkte ausgetauscht, bis ein lokales Minimum für das Optimalitätskriterium erreicht

wird. Für D-optimale Designs wurden erweiterte Austauschalgorithmen entwickelt, wie z.B. der Fedorov-Algorithmus [38], der DETMAX-Algorithmus [107] sowie der k -Exchange-Algorithmus [75], deren Lösung jedoch auch vom initialen Messplan \mathbf{M}_0 abhängt (lokales Minimum). Eine Möglichkeit zur Bestimmung eines globalen Minimums besteht darin, den Austauschalgorithmus in verschiedenen Startmessplänen \mathbf{M}_0 zu starten und das beste Ergebnis zu verwenden. Alternativ hat Poland einen genetischen Algorithmus für D-optimale Versuchspläne entwickelt, welcher das globale Minimum bestimmt [123].

Eine weitere Form der modellbasierten Messpläne sind sogenannte Netzbereichs-Messpläne. Dabei erfolgt die Messpunktverteilung nicht über den gesamten Stellgrößenraum, sondern innerhalb der jeweiligen Partitionierung eines Netzmodells [53, 71, 58]. In jedem lokalen Teilbereich wird dabei ein eigener klassischer oder raumfüllender Messplan erstellt, wodurch eine Anpassung des Messplans an die vom Netzmodell abgebildete Nichtlinearität erfolgt.

Anschauliche Beispiele sowie Überblicke zu verschiedenen Arten von Messplänen sind in [162, 66, 68] zu finden.

Messpunktreihenfolge

Bei der dynamischen Vermessung spielt nicht nur die Verteilung der Messpunkte im Eingangsraum eine Rolle, sondern auch die Sortierung der Messpunkte. Zimmerschied [173] empfiehlt eine mathematisch motivierte Sortierung, bei der je nach Prozess eine zufällige Verteilung der Amplituden mit entweder konstanter oder gegenläufiger dreieckiger Wahrscheinlichkeitsdichte erfolgt. Dadurch kann entweder die (dynamische) Raumabdeckung oder die auftretende mittlere Sprunghöhe maximiert werden. Weitere Möglichkeiten sind die Sortierung nach minimalem, maximalem oder mittlerem Messpunktabstand sowie die physikalisch motivierte temperaturoptimale Sortierung [138]. Für eine stationären Vermessung sind dabei die Sortierung nach minimalem Punktabstand oder die temperaturoptimale Sortierung am geeignetsten, da hierbei die Minimierung der thermischen Einschwingzeit die notwendige Messdauer reduziert. Bei der dynamischen Vermessung ist dagegen eine ausgeprägte Anregung erwünscht, wofür sich die Sortierung nach mittlerem oder maximalem Messpunktabstand eignen.

4.1.4 Anregungssignale zur Vermessung

Zusätzlich zur Verteilung und Reihenfolge der Messpunkte muss im Fall der Vermessung eines dynamischen Modells ein Signal zur geeigneten Anregung der Prozessdynamik ausgewählt werden. Die Amplituden der Signale werden direkt mit den sich aus dem Messplan \mathbf{M} ergebenden Stellgrößenkombinationen moduliert, ergo sind sie fest vorgegeben. Die Signaldauer und das angeregte Frequenzspektrum stellen weitere Kriterien zur Auswahl eines geeigneten Signals dar.

Sprung

Sprungsignale stellen die einfachste Signalart dar. Zudem sind sie infolge der freien Wahl der Signalamplitude sehr gut an den Prozess anpassbar. Dabei ist zu beachten, dass ein idealer Sprung technisch nicht realisierbar ist. Vielmehr ist jede sprungförmige Verstellung eines Aktors eine schnelle rampenförmige Verstellung. Unter der Annahme der Identifikation mit Sprüngen ergibt die Fehlerbetrachtung der Spektren $|u_S(i\omega)|$ von einem Sprung und $|u_R(i\omega)|$ von einer Rampe, dass die zulässige Rampenanstiegszeit t_R bei einem erlaubten Fehler von 1 % bestimmt wird durch [70]:

$$\kappa = \frac{|u_R(i\omega)|}{|u_S(i\omega)|} = \frac{\sin \frac{\omega t_R}{2}}{\frac{\omega t_R}{2}} \geq 0.99 \Rightarrow t_R \leq \frac{0.5}{\omega_{\max}}. \quad (4.12)$$

Dabei stellt ω_{\max} die höchste für die Identifikation relevante Frequenz dar. Unter der Bedingung (4.12) kann das real rampenförmige Signal bei der Identifikation als Sprung betrachtet werden. Generell sollte die Verstellung bei der Identifikation mit Sprüngen möglichst schnell sein, damit der Fehler durch die Annahme eines idealen Sprunges gering ist.

Sprungsequenz

Eine Sprungsequenz ergibt sich durch die Aneinanderreihung von verschiedenen Sprüngen, welche sich in der Zeit zwischen den Sprüngen und der Amplitude unterscheiden. Wird die Dauer von einem Sprung zum Folgenden durch ein rückgekoppeltes Schieberegister [70] erzeugt, entsteht ein DRBS (Diskretes-Rausch-Binär-Signal [172]). Dabei findet ein Vorzeichenwechsel zu diskreten Zeitpunkten $k\lambda$, $k = 1, 2, 3, \dots$ statt. Hierbei ist λ_{\min} die Länge eines Zeitintervalls (Taktzeit). Die Signallänge ergibt sich mit der Anzahl N an Zeitintervallen zu $T_{p,DRBS} = N\lambda_{\min}$. Die Anzahl N wiederum ergibt sich dabei aus der Stufenanzahl der Schieberegister n zu $N = 2^n - 1$.

Über die Vorgabe von (Nicht-)Schaltwahrscheinlichkeiten kann das Rausch-Binär-Signal auch erzeugt werden. Das DRBS hat eine konstante Schaltwahrscheinlichkeit von 50 %, so dass ein Sprung relativ häufig auftritt. Schaltet zu jedem Zeitschritt das Signal zufällig mit einer Wahrscheinlichkeit $p_{NS} \neq 50\%$ auf den nächsten Zustand, entsteht ein GRBS [172, 70]. Durch die frei vorzugebende (Nicht-)Schaltwahrscheinlichkeit p_{NS} wird eine flexible Anpassung an den zu identifizierenden Prozess ermöglicht.

Durch die periodische Wiederholung eines DRBS entsteht ein PRBS (Pseudo-Rausch-Binär-Signal [70]), welches im Gegensatz zum DRBS ein reproduzierbares deterministisches Signal ist. Durch die Erhöhung der Taktzeit werden tiefe Frequenzen stärker angeregt. Für $\lambda_{\min} = T_0$ und große N entspricht das PRBS einem weißen Rauschen, welches zur Identifikation von linearen Prozessen geeignet ist.

Soll ein nichtlinearer Prozess identifiziert werden, müssen die binären Amplituden (0, 1) der Sprünge zusätzlich mit den sich aus den Messplänen ergebenden Amplituden moduliert werden. Hieraus entstehen amplitudenmodulierte Anregungssignale (siehe [70], z.B. ADRBS oder APRBS).

Durch das Orthogonalisieren des APRBS kann ein unkorreliertes Testsignal (APRMS, Amplitudenmoduliertes Pseudo-Rausch-Mehrstufen-Signal [18, 137]) erzeugt werden. Die Signallänge des APRMS ist dabei direkt proportional zu der Anzahl an Stellgrößen und Amplitudenstufen.

Rampe

Die Rampe stellt ein dynamisches Anregungssignal dar, mit dem eine quasistatische Vermessung durchgeführt werden kann. Eine quasistatische Vermessung ist eine Mischform aus der statischen und dynamischen Vermessung. Dabei werden die Stellgrößen mit langsamen Rampen verstellt, so dass nur eine schwache dynamische Anregung ohne Einschwingen erfolgt. Die Modellbildung erfolgt rein stationär unter Eliminierung des dynamischen Einflusses. Diese kann durch eine bidirektionale Messung [144] erfolgen oder durch eine Kompensation mit einem dynamischen Prozess- und Messkettenmodell [167].

Sinus-, Chirp- und Multisinussignal

Eine weitere Möglichkeit zur Anregung von Prozessen sind Sinus- bzw. Chirp-Signale [12]. Der Zeitverlauf dieser Signale ergibt sich aus

$$u_{\text{Sin}}(t) = u_0 \sin(2\pi f_0 t) \text{ bzw. } u_{\text{Chirp}}(t) = u_0 \sin(2\pi f(t)t). \quad (4.13)$$

Das Sinussignal regt dabei nur eine feste Frequenz f an, während beim Chirp-Signal durch Vorgabe einer zeitabhängigen Frequenz $f(t)$ ein Frequenzband angeregt werden kann. Die langsamere Verstellung im Vergleich zu Sprüngen sowie die Vorgabe eines Frequenzbandes machen Chirp-Signale zu zielgerichtet an den Prozess anpassbaren Anregungssignalen, welche gut praktisch umsetzbar sind [12], jedoch lange Messzeiten erforderlich machen.

Durch additive Überlagerung von Sinussignalen mit verschiedenen Frequenzen zu einem Multisinussignal können wie beim Chirp Signal verschiedene Frequenzen gleichzeitig angeregt werden. Allerdings verstärkt eine additive Überlagerung die Amplitude über ein zulässiges Maß. Dies kann wiederum verhindert werden, indem die Überlagerung additiv mit einer Phasenverschiebung erfolgt [139]:

$$u_{\text{Multisin}}(t) = \sum_{n=1}^N u_0 \sin(2\pi f_n t + \varphi_n) \text{ mit } \varphi_n = -\frac{n(n+1)\pi}{N} \quad (4.14)$$

Wird die Überlagerung für Kreisfrequenzen im Bereich $0 \frac{\text{rad}}{\text{s}} < \omega_n \leq \pi \frac{\text{rad}}{\text{s}}$ bzw. Frequenzen im Bereich $0 \text{ Hz} < f_n \leq 0.5 \text{ Hz}$ durchgeführt, entsteht ein SoS-Signal (sum of sinusoids) [150]. Dieses ist den pseudostochastischen Signalen sehr ähnlich, da es eine starke Anregung der niedrigen Frequenzen bis $0.5 \frac{1}{\text{s}}$ [157] aufweist. Weitere Mehrfrequenz-Testsignale sind in [70] beschrieben.

Spektrumoptimale Signale

Spektrumoptimale Signale werden nach dem Kriterium der größten Übereinstimmung mit einem gewünschten Frequenzspektrum entworfen. Hierzu zählen die Multilevel Multiharmonic

Signale (MLMH) [103, 46]. Sie stellen sicher, dass die Peak-to-Peak Amplitude eingehalten wird bei maximierter Leistung im gewünschten Frequenzband. Vorteil hierbei ist, dass im Vergleich zu SoS-Signalen unerwünschte Frequenzen kaum angeregt werden, die gewünschten dagegen stärker. Dadurch ist eine bessere Anregung des vorgegebenen Frequenzbandes möglich.

4.1.5 Verfahren zur modellbasierten Optimierung

Die modellbasierte Optimierung stellt ein Anwendungsgebiet der Vermessung dar. Bei der Offline-Vermessung wird dabei nach Abschluss der Modellbildung eine stationäre Optimierung von Sollwerten für Steuerungen und Regelungen des Motorsteuergeräts durchgeführt. Die Stellgrößen des Verbrennungsmotors entsprechen dabei den zu optimierenden Variablen, während die optimierte Zielgröße je nach Anwendung der Kraftstoffverbrauch, das Motormoment, die Motoremissionen oder geeignete Kombinationen derer sind. Diese modellbasierte Optimierung der Applikation stellt aufgrund der heutzutage hohen Komplexität einen wesentlichen Schritt der Verbrennungsmotorentwicklung dar.

Die neu entwickelte Online-Methodik zur Versuchsplanung integriert in die Messpläne zusätzlich die bereits optimierten Stellgrößen, wodurch das Optimum der Zielfunktion automatisch im iterativen Verlauf mit vermessen wird. Die Verfahren der modellbasierten Optimierung werden allgemein nach dem eingesetzten Optimierungsalgorithmus sowie der Art der Optimierungszielgröße unterschieden.

Optimierungsalgorithmus

Aus der Literatur bekannte numerische Algorithmen sind gradientenfreie Verfahren, gradientenbasierte Verfahren erster und zweiter Ordnung sowie metaheuristische Algorithmen. Letztere sind globale Optimierer und können auf Kosten einer erhöhten Rechenzeit das globale Optimum einer Zielfunktion bestimmen. Die beiden zuerst genannten Verfahren sind dagegen lokale Optimierer.

Gradientenfreie Optimierungsverfahren (z.B. *simplex* Algorithmus [118]) bestimmen ausgehend von initialen Startwerten durch die iterative Auswertung der Zielfunktion die optimalen Größen. Nachteil der gradientenfreien Verfahren ist die in der Regel langsame Konvergenz sowie die Gefahr, nur ein lokales Optimum zu finden. Gradientenbasierte Verfahren (z.B. *steepest descent* Algorithmus [42]) nutzen zur Richtungsbestimmung die Ableitung der Zielfunktion. Dadurch konvergieren diese schneller als gradientenfreie Verfahren, können aber auch in einem lokalen Optimum hängen bleiben. Bei beiden Verfahrensklassen kann das Risiko des Hängenbleibens in einem lokalen Optimum durch die Initialisierung an mehreren Stellen im Versuchsraum reduziert werden.

Metaheuristische Algorithmen sind Optimierungsverfahren, welche sich nach definierten Regeln aus einem lokalen Optimum wieder heraus iterieren und dadurch auch ein globales Optimum bestimmen können. Dem gegenüber steht jedoch die zeitintensive Bestimmung der optimierten Größen. Zu den bekannten Verfahren zählen *evolutionäre* Algorithmen [145], die *simu-*

lated annealing Optimierung [82] sowie die *particle swarm* Optimierung [79]. Erstere werden vor allem bei Mehrziel-Optimierungen eingesetzt, bei denen mehrere Zielgrößen gewichtet zueinander optimiert werden. Verfahren wie der NSGA-II Algorithmus [27, 14] bestimmen durch eine geeignete Sortierung, Auswahl und Mutation der optimierten Größen direkt die sich ergebende optimale Pareto-Front der gewichteten Zielgrößen.

Optimierungszielfunktion

Die Gütefunktionen der modellbasierten Applikation, welche die Optimierungszielgrößen enthalten, werden im Rahmen der Optimierung von Verbrennungsmotoren unterschieden zwischen lokalen versus globalen sowie statischen versus dynamischen Gütefunktionen. An dieser Stelle sind diese Unterschiede allgemein beschrieben. Konkrete Beispiele für die Zielfunktionen im Falle der Emissionsoptimierung folgen in Abschnitt 4.3.4.

Die Optimierung mit einer lokalen Zielfunktion f_o , welche die von den Stellgrößen abhängigen Zielgrößen geeignet kombiniert, innerhalb eines Variationsraumes VR_o erfolgt gemäß folgender Gleichung:

$$\mathbf{u}_{\text{opt,lokal},o} = \arg \min_{\mathbf{u}_o \in VR_o} \left(f_o(\mathbf{u}_o) \right) \quad (4.15)$$

Dabei ist zu berücksichtigen, dass die lokale Zielfunktion sowohl mit einem lokalen als auch einem globalen Optimierer minimiert werden kann. Sequenz [146] erweitert die Zielgröße additiv um zwei weitere Terme:

$$\mathbf{u}_{\text{opt,lokal},o} = \arg \min_{\mathbf{u}_o \in VR_o} \left(f_o \pm \sqrt{\mathbf{x}^T \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}} + \left| \frac{\delta f_o}{\delta \mathbf{u}_o} \right| \sigma_{\text{sensor},\mathbf{u}} \right) \quad (4.16)$$

Der erste Term stellt dabei die zu minimierende Zielfunktion dar. Durch den zweiten Term werden die Modellunsicherheiten bei einer Schätzung mit der Methode der kleinsten Quadrate berücksichtigt. Der letzte Term erweitert die Optimierung um eine Serienstreuung, indem für die optimierten Stellgrößen das Produkt aus Sensitivität (Steigung im betrachteten Punkt) und maximaler Streuung (Varianz der Sensoren der Stellgrößen) beachtet wird.

Die globale Zielfunktion optimiert im Gegensatz zur lokalen Zielfunktion mehrere Arbeitspunkte bzw. innerhalb mehrerer Variationsräume gleichzeitig. Dabei wird jeder Arbeitspunkt $o = 1, \dots, O$, gewichtet mit der Verweildauer w_o im betrachteten Emissionszyklus, zur Zielfunktion summiert und die optimalen Stellgrößen für jeden Arbeitspunkt bestimmt:

$$\mathbf{u}_{\text{opt,global}} = \arg \min_{\mathbf{u} \in VR} \left(f(\mathbf{u}) \right) = \arg \min_{\mathbf{u}_o \in VR_o} \left(\sum_{o=1}^O w_o \cdot f_o(\mathbf{u}_o) \right) \quad (4.17)$$

4.2 Anregungssignal zur Identifikation von stationären und dynamischen Modellen

Zur kombinierten Bestimmung des statischen und dynamischen Prozessverhaltens in verschiedenen Messpunkten wird ein deterministisches Testsignal entworfen. Durch die adäquate Kom-

bination einer dynamischen Anregung und eines Einschwingvorgangs ins stationäre Prozessverhalten kann das dynamische Modell des Prozesses identifiziert werden. Dieses weist nach Überführung in ein statisches Modell (siehe Abschnitt 5.1.3), auch im statischen, eine hohe Güte auf.

Dieses neu vorgeschlagene deterministische dynamische Anregungssignal wird gebildet, indem ein Rechteck- und ein Sprungsignal zu einem *deterministischen binär Signal* (DBS) kombiniert werden. Bild 4.4 a) zeigt die Kombination von einem Rechteckimpuls mit einer Sprungfunktion zu einem DBS. Die Wahl der Taktzeit λ_{\min} , welche dem am kürzesten auftretenden Impuls entspricht, ergibt sich entsprechend den Regeln eines (A)PRBS [120, 70] aus einem vielfachen der Abtastzeit T_0 (μT_0 mit $\mu = 1, 2, \dots$). Die maximale Taktzeit λ_{\max} und somit der längste auftretende Impuls, ergibt sich beim APRBS aus der Anzahl an Schieberegisterstufen n zu $\lambda_{\max} = n \cdot \lambda_{\min}$. Im Gegensatz zum APRBS kann die maximale Taktzeit beim DBS frei gewählt werden. Mit der Festlegung von λ_{\min} und λ_{\max} zu

$$\begin{aligned}\lambda_{\min} &= \frac{T_E}{3} \quad \text{und} \\ \lambda_{\max} &= 4 \cdot \lambda_{\min} = \frac{4T_E}{3}\end{aligned}\tag{4.18}$$

ergibt sich die Periodenlänge des DBS gemäß Bild 4.4 a) zu

$$T_{p,\text{DBS}} = (1 + 3 + 4) \cdot \lambda_{\min} = 8 \cdot \lambda_{\min}.\tag{4.19}$$

Durch diese Annahmen kann erreicht werden, dass der Prozess sowohl nach dem ersten Rechteckimpuls als auch nach Ablauf der gesamten Periodendauer $T_{p,\text{DBS}}$ in guter Näherung stationär ist, falls T_E die 95 % Einschwingzeit des (nichtlinearen) Prozesses ist. Infolgedessen kann mit dem sich ergebenden 1-3-4 Signal sowohl das dynamische als auch das statische Prozessverhalten angeregt bzw. erfasst werden.

Die Amplitudenmodulation für nichtlineare Prozesse ergibt sich (im Gegensatz zum APRBS, welches durch Zuordnung einer neuen Amplitude bei jedem Signalsprung eines PRBS gebildet wird) durch Aneinanderreihung mehrerer DBS-Impulsfolgen (Bild 4.4 a) mit Amplituden bzw. Sprüngen zu neuen Vermessungspunkten, welche sich aus der statistischen Versuchsplanung ergeben (Bild 4.4 b). Dies ergibt dann ein amplitudenmoduliertes DBS (ADBS), welches exemplarisch in Bild 4.4 c) für 5 Amplituden dargestellt ist.

In Bild 4.5 sind jeweils eine Periode eines Rechtecksignals mit $T = 15$ s, eines PRBS mit $\lambda_{\min} = 1$ s und $n = 4$, eines DBS mit $\lambda_{\min} = 1.875$ s sowie in b) die sich ergebenden normierten Amplitudendichten dargestellt. Durch den Vergleich von Signalen mit gleichen Periodenlängen und Amplituden sowie den Einsatz der bezogenen Kreisfrequenz

$$\omega^* = \frac{\omega T}{2\pi}\tag{4.20}$$

sind die Amplitudendichten unabhängig von der Periodenlänge T und Amplitude u . Hieraus folgt, dass sich die Fourier-Transformierte ausschließlich durch die Signalform und nicht die Signalparametrierung ergibt [70]. Zusätzlich sind alle Amplitudendichten aus Anschauungsgründen auf die maximale Fourier-Amplitude des Rechtecksignals normiert.

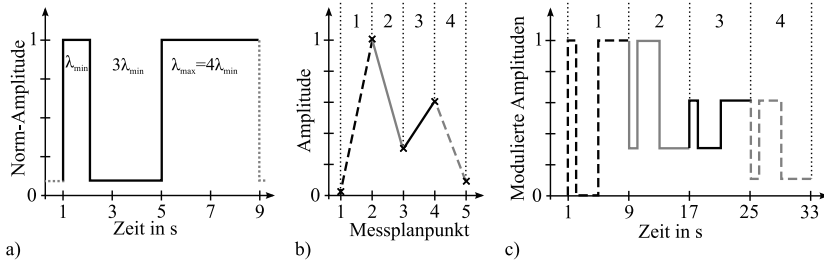


Bild 4.4: Erzeugung eines DBS (1-3-4 Signal) zur dynamischen Anregung von Prozessen. Durch Kombination eines Rechteckimpulses mit einem Sprung entsteht das DBS (a), welches zusammen mit Amplituden aus der statistischen Versuchsplanung (b) zu einem ADBS für die Anregung von nichtlinearen dynamischen Prozessen (c) moduliert werden kann.

Das Rechtecksignal hat aufgrund der längsten Dauer ohne Signalwechsel zwangsläufig immer die stärkste Anregung bei tiefen Frequenzen $\omega^* < 0.7$, das PRBS aufgrund der vielen Signalwechsel entsprechend die schwächste Anregung. Im Bereich $0.7 < \omega^* < 2.7$ regt das DBS einen Prozess überwiegend am stärksten an. Das PRBS hat gegenüber den zwei anderen Signalen erst ab Frequenzen $\omega^* > 2.7$ Vorteile.

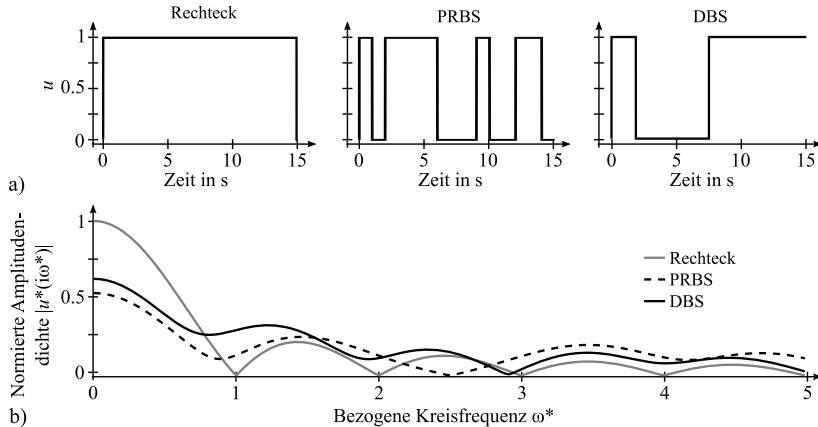


Bild 4.5: Vergleich der Amplitudenspektren von jeweils einer Periode eines Rechtecksignals, einem PRBS sowie einem DBS zur Prozessanregung. Die Amplitudendichte des DBS ist im tiefen Frequenzbereich größer als die des PRBS sowie im mittleren Bereich $0.7 < \omega^* < 2.7$ überwiegend am größten aller drei Signale.

4.2.1 Vergleich der Signale zur dynamischen Prozessanregung

Zum Vergleich der verschiedenen Anregungssignale wird nur die *lineare* dynamische Übertragungsfunktion des insgesamt nichtlinearen Hammerstein-Prozesses aus Gleichung (D.1) für eine Eingangs- und eine Ausgangsgröße simuliert und identifiziert. Hierdurch kann die Güte der Identifikation der Dynamik unabhängig von der Güte der Identifikation einer möglichen Nichtlinearität bestimmt werden. Die Einschwingzeit ergibt sich für diesen Prozess zu ca. $T_E = 12$ s und die Abtastzeit wird zu $T_0 = 0.3$ s gewählt. Die Übertragungsfunktion ändert sich im Gegensatz zu (D.1) aufgrund der Abtastzeit zu:

$$y(z) = \frac{0.1616z + 0.149}{z^2 - 1.476z + 0.7866} \tilde{u}(z) \quad (4.21)$$

Die dynamisch anzuregenden Messpunkte (Amplituden) zur Identifikation des linearen dynamischen Prozesses werden als 0 und 1 gewählt, um die Anregung zu maximieren, wodurch sich die Identifikation der kleinsten Fehler bei maximaler Information ergibt [89]. Angeregt wird der Prozess mit einem DBS mit $\lambda_{\min} = 2$ s, woraus sich eine Signallänge von $T_p = 16$ s berechnet. Zusätzlich wird der Prozess mit einem PRBS mit der Schieberegisterlänge $n = 4$ identifiziert, um die gleiche maximale Taktzeit $\lambda_{\max} = 8$ s wie das DBS zu erhalten. Das PRBS hat dabei eine Länge von Signallänge von $T_p = 30$ s und ist somit annähernd doppelt so lang wie das DBS. Die Parameterschätzung des Modells erfolgt unter Annahme einer festen Modellstruktur in Ausgangsfehleranordnung:

$$\Psi = \begin{bmatrix} \tilde{u}_1(k-1) & \tilde{u}_1(k-2) & -\hat{y}_1(k-1) & -\hat{y}_1(k-2) \\ \tilde{u}_2(k-1) & \tilde{u}_2(k-2) & -\hat{y}_2(k-1) & -\hat{y}_2(k-2) \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{u}_N(k-1) & \tilde{u}_N(k-2) & -\hat{y}_N(k-1) & -\hat{y}_N(k-2) \end{bmatrix} \quad (4.22)$$

Die resultierenden Modelle werden mit einem Signal bestehend aus Amplituden eines Raster-Messplans mit 20 Messpunkten, zufälligen Pulsweiten im Bereich von 2...16 s sowie zufällig verteilten rampen- und sprungförmigen Anregungen validiert. Die daraus folgenden Signalverläufe für das Training und die Validierung sind in Bild 4.6 für das DBS und die PRBS dargestellt. Hierbei sind in a) die Signalverläufe der Stellgröße für beide Anregungen sowie die zugehörigen, auf die Signalflächen normierten, bezogenen Frequenzspektren dargestellt. Das PRBS, mit gleicher maximaler Taktlänge wie das DBS, zeigt eine ähnlich starke Anregung von tiefen Frequenzen, wobei jedoch die mittleren Frequenzen schwächer angeregt werden. Durch den Einsatz eines DBS-Signals wird somit ein breiteres Frequenzspektrum bei kürzerer Signallänge angeregt, als es mit einem PRBS der Fall ist. In Bild 4.6 b) ist der Verlauf der Eingangsgröße des insgesamt 180 s langen Validierungssignals dargestellt sowie die sich für die beiden Modelle ergebenden Ausgangsgrößenverläufe. Beide Anregungssignale zeigen eine sehr gute Übereinstimmung in der Validierung. Die Dynamik und das stationäre Verhalten werden sehr gut identifiziert. Die resultierenden Übertragungsfunktionen mit den geschätzten Parametern folgen zu

$$\begin{aligned} y_{\text{PRBS}}(z) &= \frac{0.3843z - 0.0894}{z^2 - 1.4642z + 0.7659} \tilde{u}(z) \\ y_{\text{DBS}}(z) &= \frac{0.4153z - 0.1074}{z^2 - 1.4625z + 0.7761} \tilde{u}(z) \end{aligned} \quad (4.23)$$

mit RMSE-Validierungsgüten der Modelle von

$$\begin{aligned} J_{\text{RMSE,PRBS}}(z) &= 0.0583 \\ J_{\text{RMSE,DBS}}(z) &= 0.0407. \end{aligned} \quad (4.24)$$

Der Vergleich der geschätzten Parameter sowie der Validierungsgüten macht deutlich, dass beide vorgestellten Anregungssignale für eine Identifikation der Dynamik des Prozesses sehr gut geeignet sind, wobei die Schätzung mit dem DBS eine geringfügig bessere Güte bei gleichzeitig einer ca. 47 % kürzeren Signallänge aufweist.

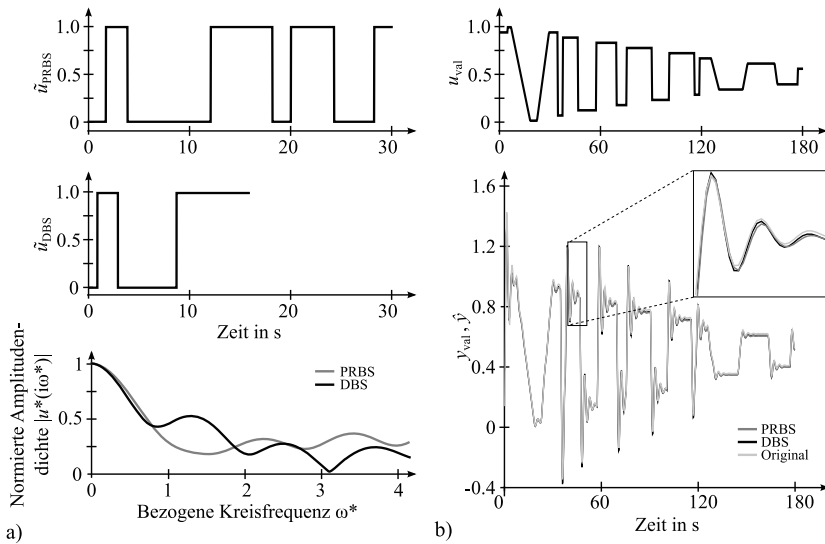


Bild 4.6: Vergleich der Identifikation eines dynamischen Beispielprozesses für die zwei Anregungssignale $\text{PRBS}_{n=4}$ und DBS. Dargestellt sind jeweils der zeitliche Verlauf und das Frequenzspektrum der Stellgrößen für das Training (a) sowie der Verlauf der Ein- und modellierten Ausgangsgrößen der Validierung (b).

Zur Bewertung der Güte der Identifikation eines *nichtlinearen* Prozesses, wird in einem zweiten Beispiel der vollständige nichtlineare Hammerstein-Prozess aus Gleichung (D.1) identifiziert. Als Messplan zur Erfassung der statischen Nichtlinearität des dynamischen Prozesses kommt ein Sobol-Messplan zum Einsatz. Zur Identifikation des dynamischen Prozessverhaltens werden die sich aus dem Messplan ergebenden Amplituden auf die beiden Anregungssignale moduliert.

Angeregt wird der Prozess mit jeweils einem ADBS mit $\lambda_{\min} = 4$ s und $\lambda_{\max} = 12$ s und einem APRBS mit Schieberegisterlänge $n = 7$, wodurch sich 64 Impulse mit einer maximalen Impulsdauer von $\lambda_{\max} = 28$ s ergeben. Die sich daraus ergebenden Signallängen betragen 504 s bzw. 480 s. Die Modellbildung erfolgt wieder mit einer LS-Schätzung der Parameter für eine

feste Modellstruktur in Gleichungsfehleranordnung:

$$\Psi = \begin{bmatrix} \tilde{u}_{1,1}(k-1) & \tilde{u}_{1,1}(k-2) & \tilde{u}_{1,1}^2(k-1) & \tilde{u}_{1,1}^2(k-2) & \tilde{u}_{2,1}^2(k-1) & \dots \\ \tilde{u}_{2,1}(k-1) & \tilde{u}_{2,1}(k-2) & \tilde{u}_{2,1}^2(k-1) & \tilde{u}_{2,1}^2(k-2) & \tilde{u}_{2,2}^2(k-1) & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots \\ \tilde{u}_{N,1}(k-1) & \tilde{u}_{N,1}(k-2) & \tilde{u}_{N,1}^2(k-1) & \tilde{u}_{N,1}^2(k-2) & \tilde{u}_{N,2}^2(k-1) & \dots \end{bmatrix} \quad (4.25)$$

$$\begin{bmatrix} \tilde{u}_{2,1}^2(k-2) & \tilde{u}_{2,1}^3(k-1) & \tilde{u}_{2,1}^3(k-2) & -y_1(k-1) & -y_1(k-2) \\ \tilde{u}_{2,2}^2(k-2) & \tilde{u}_{2,2}^3(k-1) & \tilde{u}_{2,2}^3(k-2) & -y_2(k-1) & -y_2(k-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{u}_{N,2}^2(k-2) & \tilde{u}_{N,2}^3(k-1) & \tilde{u}_{N,2}^3(k-2) & -y_N(k-1) & -y_N(k-2) \end{bmatrix}$$

Nach erfolgter Identifikation werden die resultierenden Modelle mit einem Signal bestehend aus 81 rasterförmig verteilten Amplituden ($L = 9$) und zufälligen Pulsweiten zwischen 2 s und 16 s validiert. Die sich ergebenden Zeitverläufe der zum Training verwendeten zwei Anregungssignale sowie ein Ausschnitt aus dem 750 s langen Ein- und Ausgangssignal für die Validierung der Identifikation sind in Bild 4.7 a) und b) dargestellt. Beide Anregungssignale zeigen eine gute Übereinstimmung in der Validierung. Die Dynamik und das stationäre Verhalten werden sehr gut identifiziert. Die RMSE-Validierungsgüten der Modelle folgen zu

$$J_{\text{RMSE,APRBS},n=7}(z) = 0.8337$$

$$J_{\text{RMSE,ADBS}}(z) = 0.7173. \quad (4.26)$$

Zusammenfassend kann festgehalten werden, dass beide Signale für eine Identifikation der Dynamik und Nichtlinearität eines Prozesses gut geeignet sind. Mit dem ADBS lassen sich dabei nichtlineare Prozesse flexibler anregen, wodurch eine höhere Modellqualität erreicht wird bzw. eine (vom Prozess abhängige) kürzere Signallänge im Vergleich zum APRBS zum Erreichen der selben Modellqualität ausreichend ist.

4.3 Aktiv lernende Messpläne

Die optimale Platzierung von Messpunkten basierend auf Informationen aus Messungen wird als aktiv lernende Versuchsplanung [123] bzw. Online-Versuchsplanung bezeichnet.

Wie in Abschnitt 4.1.2 beschrieben, erfolgt die Effektermittlung (Regressorauswahl), integriert in der Gesamtmethodik bei der Modellbildung 5.4.5 auf die Weise, dass die relevanten Einflussgrößen sowie Effekte direkt aus der Modellbildung folgen und für die Versuchsplanung übernommen werden können. Die Versuchsplanung im Rahmen der Online-Methodik reduziert sich somit auf die Bestimmung der Messpunktverteilung bzw. Messpläne.

Die in Abschnitt 4.1.3 vorgestellten Methoden zur Erstellung von Messplänen sind entweder vom Prozess unabhängig oder basieren auf einer Modellannahme. Keines der Kriterien nutzt gemessene Werte y zur Erstellung des Messplans und somit kann aus Messungen kein Informationsgewinn zur Erzeugung der Messpläne erfolgen. Eine Online-Versuchsplanung nutzt die

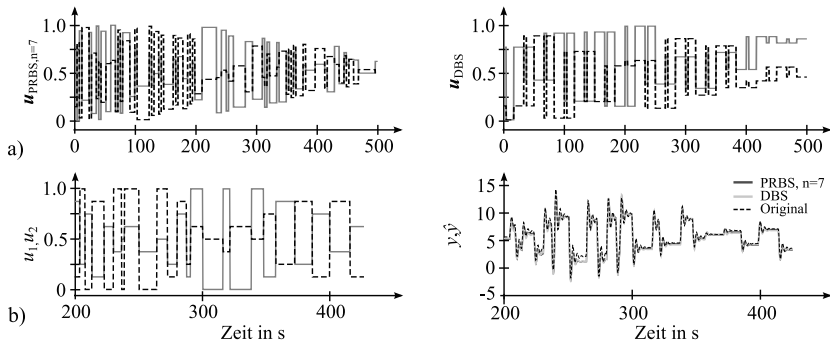


Bild 4.7: Vergleich der Identifikation eines dynamischen nichtlinearen Beispielprozesses für zwei Anregungssignale PRBS_{n=7} und DBS. Dargestellt sind beispielhaft a) für das Training die Eingangsgrößen für jedes Anregungssignal und b) für die Validierung ein 230 s Ausschnitt der Ein- und modellierten Ausgangsgrößen.

iterativ gewonnenen Vermessungsinformationen dazu, den Messplan mithilfe von statistischen und modellbasierten Methoden an den realen Prozess anzupassen. Diese zielgerichtete Anpassung führt zu einer systematischen, fortlaufenden Erhöhung der sich aus neuen Messpunkten ergebenden Modellgüte.

4.3.1 Gesamtstrategie zur Online-Messplanerweiterung

Die Gesamtstrategie zur aktiv lernenden Online-Messplanerweiterung besteht aus mehreren Schritten. Sie ist in Bild 4.8 dargestellt. Sie unterteilt sich dabei in die iterative Anpassung des Stellbereichs sowie die iterative Erweiterung des Messplans.

Bei der Adaption des Stellbereichs wird, unter Einbezug der Messgrößenvarianz, der Bereich bestimmt, in welchem das aktuelle Modell den größten Fehler aufweist (siehe Abschnitt 4.3.2). Im Anschluss daran werden mithilfe der Erweiterung des Messplans innerhalb des adaptierten Bereiches basierend auf dem raumfüllenden (DTS-Verfahren, Abschnitt 4.3.3) sowie modellbasierten Verfahren (MQI, MQD, MPC und MO-Verfahren, Abschnitt 4.3.4) neu zu vermessende Messpunktkandidaten zielgerichtet bestimmt. Diese ergeben sich dabei direkt aus den aktuell vorhandenen Prozessinformationen.

Dieses Vorgehen wird iterativ wiederholt. Durch die gezielte Messpunktplatzierung erfolgt die Modellverbesserung gezielt in den Bereichen, in denen das Modell am ungenausten und somit die Vermessung am wirksamsten ist.

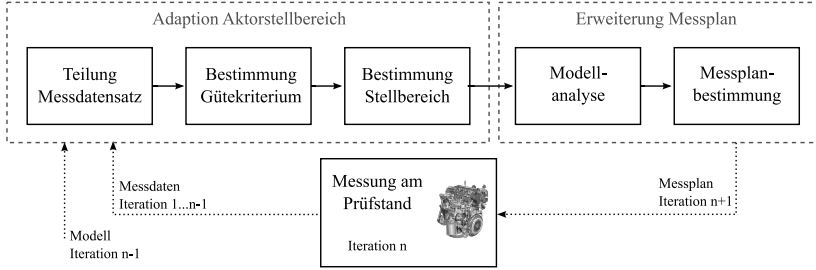


Bild 4.8: Darstellung der Gesamtstrategie zur aktiv lernenden Online-Messplanerweiterung bestehend aus der Adaption des Stellbereichs und Erweiterung des Messplans in der Iteration n basierend auf den Messdaten der Iterationen $1 \dots n-1$ und dem Modell der Iteration $n-1$.

4.3.2 Iterative Stellbereichsanpassung

Ausgehend von den iterativ gemessenen Daten und dem daraus identifizierten Modell (gemäß der Methoden in Kapitel 5) wird eine Adaption der in der nächsten Iteration zu vermessenden Aktorstellbereiche durchgeführt. Dabei wird die von lokal-linearen Netzmodellen stammende Anpassung an die Nichtlinearität durch eine orthogonale Teilung des Netzes, basierend auf der lokalen Güte (LOLIMOT [120, 70]), bei der Stellbereichsanpassung genutzt. Hierfür wird der Messdatensatz, bestehend aus allen gemessenen Ein- und Ausgangsdaten, in allen Stelldimensionen nacheinander mit einem festen Teilungsverhältnis von $l = 0.5$ aufgeteilt. Die Werte $\mathbf{u}_{\text{split}}$, bei denen der Messdatensatz geteilt wird, folgen aus

$$\mathbf{u}_{\text{split}} = \mathbf{u}_{\min} + l \cdot (\mathbf{u}_{\max} - \mathbf{u}_{\min}). \quad (4.27)$$

Die resultierende Einteilung der gemessenen Daten in die sich ergebenden lokalen Modell-Regionen (LMR) ergeben sich exemplarisch für den 2-dimensionalen Fall zu (siehe Bild 4.9):

$$\begin{aligned} \text{LMR1} : \begin{pmatrix} u_{1,\min} \\ u_{2,\min} \end{pmatrix} \leq \mathbf{u}_{\text{LMR1}} \leq \begin{pmatrix} u_{1,\max} \\ u_{2,\text{split}} \end{pmatrix} & \quad \text{LMR2} : \begin{pmatrix} u_{1,\min} \\ u_{2,\text{split}} \end{pmatrix} \leq \mathbf{u}_{\text{LMR2}} \leq \begin{pmatrix} u_{1,\max} \\ u_{2,\max} \end{pmatrix} \\ \text{LMR3} : \begin{pmatrix} u_{1,\text{split}} \\ u_{2,\min} \end{pmatrix} \leq \mathbf{u}_{\text{LMR3}} \leq \begin{pmatrix} u_{1,\max} \\ u_{2,\max} \end{pmatrix} & \quad \text{LMR4} : \begin{pmatrix} u_{1,\min} \\ u_{2,\min} \end{pmatrix} \leq \mathbf{u}_{\text{LMR4}} \leq \begin{pmatrix} u_{1,\text{split}} \\ u_{2,\max} \end{pmatrix} \end{aligned} \quad (4.28)$$

Mithilfe der sich ergebenden lokalen Datensätze $\mathbf{u}_{\text{LMR}j}$ (mit $j = 1, \dots, 4$) werden lokale Modelle identifiziert. Ausgehend von den lokalen Datensätzen und Modellen wird, basierend auf einer Kenngröße der lokalen Modellgüte $J_Q(\text{LMR}j)$ sowie einer Kenngröße der lokalen Messgrößenstandardabweichung $J_S(\text{LMR}j)$, ein resultierendes Gesamtkriterium $J_A(\text{LMR}j)$ für jeden der aktuellen Stellbereiche $\text{LMR}j$ bestimmt. Dadurch kann sowohl die Modellgenauigkeit als auch die Genauigkeit der Messgrößen in Form ihrer Standardabweichung im stationären Zustand bei der iterativen Anpassung der Stellbereiche integriert werden.

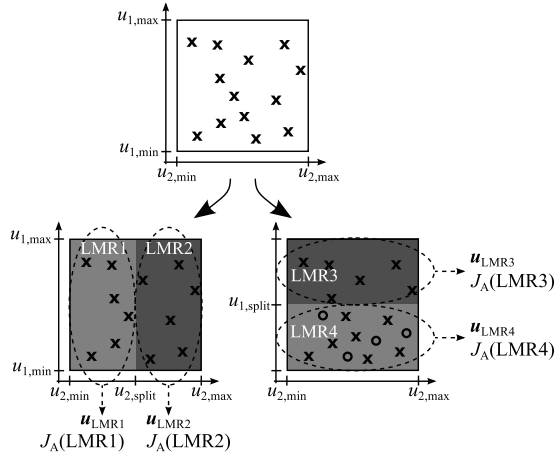


Bild 4.9: Vorgehen zur Bestimmung des nächsten zu vermessenden Bereichs. Neue Messpunkte (Kreise) werden in den nach mehrfacher Teilung schlechtesten Raum bzgl. des Gesamtkriteriums J_A gelegt (hier im Bild LMR4).

Kenngröße zur Berücksichtigung der Modellgüte

Mithilfe der lokalen Modelle in den jeweiligen Stellbereichen LMR_j werden die lokalen Modellgüten $J_{\text{NRMSEv}}(\text{LMR}_j)$ bzw. $J_{\text{NRMSEs}}(\text{LMR}_j)$ in allen j Teilbereichen bestimmt. Dies erfolgt durch die Mittelung aller im aktuellen Stellbereich LMR_j gemessenen Ausgangsgrößen $i = 1, \dots, O$ zu einem lokalen Gütekriterium:

$$J_{\text{RMSE,lokal}}(\text{LMR}_j) = \sqrt{\frac{\sum_{i=1}^O (y_i - \hat{y}_{j,i})^2}{O}} \quad (4.29)$$

Hierbei ist \hat{y} die modellierte und y die gemessene Ausgangsgröße im lokalen Stellbereich LMR_j . Für eine zielgerichtete Vermessung müssen im Rahmen der Online-Methodik neue Messdatenpunkte unter Berücksichtigung der zunehmenden Anzahl an Messdaten bestimmt werden. Hierfür wird das im lokalen Stellbereich gemittelte Fehlerkriterium zusätzlich um die Berücksichtigung des Volumens der Hyperquader bzw. der Anzahl an Messpunkten erweitert. Das Kriterium zur Berücksichtigung der Modellgüte J_Q ergibt sich durch die Multiplikation des gemittelten Kriteriums $J_{\text{RMSE,lokal}}(\text{LMR}_j)$ mit dem Verhältnis der Hypervolumen bzw. der Anzahl an Messsamples:

$$J_{\text{NRMSEv}}(\text{LMR}_j) = J_{\text{RMSE,lokal}}(\text{LMR}_j) \cdot \frac{V_{\text{LMR}}}{V_{\text{GMR}}}, \quad (4.30)$$

$$J_{\text{NRMSEs}}(\text{LMR}_j) = J_{\text{RMSE,lokal}}(\text{LMR}_j) \cdot \frac{O}{N}. \quad (4.31)$$

Durch die Multiplikation des Fehlerkriteriums mit dem Verhältnis des sich aus dem lokalen Stellbereich ergebenden Hyperquadvolumens V_{LMR} zu dem globalen Hyperquadvolumen

V_{GMR} ergibt sich ein volumen-bezogenes Kriterium (J_{NRMSE_V}). Entsprechend folgt durch Multiplikation mit dem Verhältnis der im lokalen Stellbereich vorhandenen Messpunkte O zu den insgesamt vorhandenen Messpunkten N das sample-bezogene (J_{NRMSE_S}) Fehlerkriterium.

Je nachdem welches der beiden Kriterien genutzt wird, erfolgt eine stärkere Reduktion des $J_{\text{RMSE}_{\text{lokal}}}$ -Fehlerkriteriums in kleinen lokalen Stellbereichen bzw. in lokalen Stellbereichen mit wenig Messpunkten im Vergleich zu größeren Stellbereichen bzw. Stellbereichen mit vielen Messpunkten. Bei der Adaption der Stellbereiche kann dadurch neben dem Modellfehler auch die Größe der lokalen Hyperquader bzw. die Anzahl an Messpunkten innerhalb der lokalen Hyperquader berücksichtigt werden. Dabei wird die Annahme getroffen, dass das Vermessen von weiteren Messpunkten eine Reduktion des Trainingsfehlers zur Folge hat (sinkender Varianzfehler).

Das Kriterium zur Berücksichtigung der Hypervolumen ist im Rahmen der Online-Methodik bei der Vermessung mit iterativ raumfüllenden Messplänen geeignet, welche zukünftigen Messpunkte gleichmäßig über den gesamten lokalen Stellbereich generieren:

$$J_Q(\text{LMR } j) = \frac{1}{1 + J_{\text{NRMSE}_V}(\text{LMR } j)} \quad (4.32)$$

Durch die bevorzugte Generierung von raumfüllenden Messpunkten in lokalen Stellbereichen mit großen Fehlern und großen Hypervolumen sind größere Stellbereiche von der Fehlerreduktion betroffen, wodurch eine stärkere Zunahme der Modellgüte erwartet werden kann als ohne Berücksichtigung des Volumens.

Im Gegensatz dazu wird bei der Vermessung mit modellbasierten Messplänen das sample-bezogene Kriterium eingesetzt, wodurch die Anzahl an lokalen Messpunkten berücksichtigt wird:

$$J_Q(\text{LMR } j) = \frac{1}{1 + J_{\text{RMSE}_{\text{lokal}}}(\text{LMR } j) - J_{\text{NRMSE}_S}(\text{LMR } j)} \quad (4.33)$$

Indem bevorzugt in Stellbereichen mit großen Fehlern und wenigen Messpunkten neue zu vermessende modellbasierte Messpunkte generiert werden, kann eine gezielte Erhöhung der Modellqualität erreicht werden.

Kenngröße zur Berücksichtigung der Messgrößenvarianz

Zusätzlich erfolgt eine Analyse der mittleren Standardabweichung aller relevanten Größen in den j lokalen Stellbereichen. Die relevanten Größen sind dabei die Modelleingänge x_e mit $e = 1 \dots, E$ und die Modellausgänge y_a mit $a = 1 \dots, A$ sowie als Kenngröße für die Motorlaufruhe die Standardabweichung der Zylinderdrücke $\iota = \sigma_{\Delta p_{\text{Zyl}}}$. Für jede zu bestimmende Stellgrößenkombination $o = 1, \dots, O$ kann während der stationären Vermessung online die Standardabweichung σ der jeweiligen Größe über $k = 1, \dots, K$ zu mittelnende Abtastwerte be-

stimmt werden:

$$\sigma_{x_e}(\text{LMR } j) = \frac{\sum_{o=1}^O \sqrt{\frac{1}{(K-1)} \sum_{k=1}^K (x_{e,o,k} - \bar{x}_{e,o})^2}}{O} \quad (4.34)$$

$$\sigma_{y_a}(\text{LMR } j) = \frac{\sum_{o=1}^O \sqrt{\frac{1}{(K-1)} \sum_{k=1}^K (y_{a,o,k} - \bar{y}_{a,o})^2}}{O} \quad (4.35)$$

$$\sigma_l(\text{LMR } j) = \frac{\sum_{o=1}^O \sqrt{\frac{1}{(K-1)} \sum_{k=1}^K (l_{o,k} - \bar{l}_o)^2}}{O} \quad (4.36)$$

Hierbei werden die Mittelwerte $\bar{x}_{e,o}$, $\bar{y}_{a,o}$ und \bar{l}_o der jeweiligen Größen über alle K Abtastwerte bestimmt.

Bei der Bestimmung einer Kenngröße für alle Größen müssen die unterschiedlichen physikalischen Dimensionen der jeweiligen Größe einbezogen werden, da auch der Mittelwert bzw. die Standardabweichung eine entsprechende physikalische Dimension besitzen. Dadurch können diese charakteristischen Werte nicht einfach kombiniert werden. In Bild 4.10 ist dieser Zusammenhang exemplarisch für die zwei Größen Raildruck p_{Rail} und Position des AGR-Ventils s_{HDAGR} dargestellt, unter der Annahme einer Normalverteilung. Sowohl der gemessene Mittelwert beider als auch die Standardabweichung weichen signifikant voneinander ab. Eine Kombination beider Dichtefunktionen zu einer Kenngröße ist erst möglich, wenn eine geeignete Normierung erfolgt.

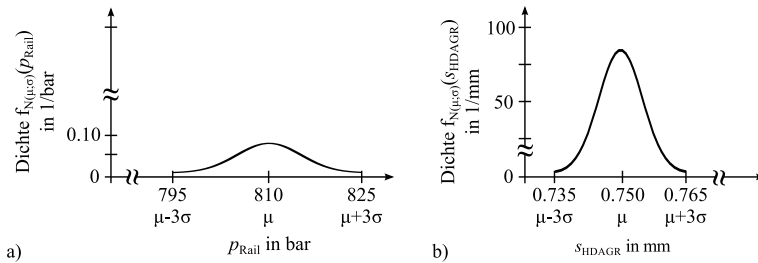


Bild 4.10: Vergleich der Dichtefunktion des Raildrucks p_{Rail} und der Position des AGR-Ventils s_{HDAGR} unter Annahme einer Normalverteilung.

Die Normierung der Messgrößenstandardabweichung erfolgt dabei mit einer an die kumulative Verteilungsfunktion der Normalverteilung angelehnten Sigmoidfunktion (siehe Bild 4.11). Sie ergibt sich spezifisch für jede physikalische Größe, sodass eine einzelne individuelle Parametrierung der Sigmoidfunktion nicht nötig ist. Bei der Verteilungsfunktion wird die Wahrscheinlichkeit einer normalverteilten Zufallsgröße auf die Weise dargestellt, dass diese kleiner als ein Mittelwert mit Standardabweichung ist. Im Gegensatz dazu stellt die Sigmoidfunktion eine Kenngröße über die Standardabweichung dar. Der x-Achsen Bereich der Sigmoidfunktion zeigt nicht die Zufallsgröße aufgespannt über Mittelwert und Standardabweichung wie in der Verteilungsfunktion, sondern allein eine angepasste Standardabweichung σ^* . Die tatsächlich

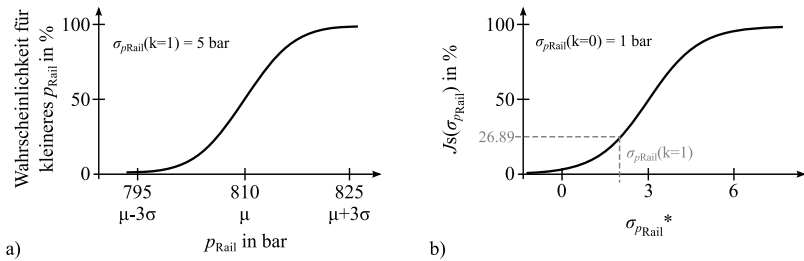


Bild 4.11: Darstellung der a) Verteilungsfunktion der Normalverteilung $N(\mu, \sigma^2)$ sowie b) der aus ihr entwickelten Sigmoidfunktion zur Gewichtung der Messgrößenstandardabweichung für den Raildruck aus Bild 4.10.

gemessene Standardabweichung $\sigma(k \neq 0)$ wird dabei auf die zu Beginn der Vermessung im sicheren Zentralpunkt gemessene Standardabweichung $\sigma(k=0)$ der Größe bezogen:

$$\sigma^* = \frac{\sigma(k \neq 0)}{\sigma(k=0)}. \quad (4.37)$$

Durch diese Transformation können alle (physikalisch unterschiedlichen) Größen kombiniert werden. Der y-Achsen Bereich der Sigmoidfunktion stellt die Kenngröße J_S dar. Er wird, ähnlich der Wahrscheinlichkeit der Verteilungsfunktion, zwischen 0% und 100% skaliert unter Berücksichtigung eines zusätzlichen x-Achsen-Offsets von 3:

$$J_S = \frac{100\%}{1 + \exp^{-(\sigma^* - 3)}}. \quad (4.38)$$

Durch die Verschiebung um den Offset steigt die Kenngröße erst bei gemessenen Standardabweichungen von $\sigma(k \neq 0) > 3\sigma(k=0)$ bzw. bezogenen Standardabweichungen $\sigma^* > 0$ signifikant an. Messgrößen mit gemessenen Standardabweichungen $\sigma(k \neq 0) > 9\sigma(k=0)$ bzw. bezogenen Standardabweichungen $\sigma^* > 6$ besitzen eine Kenngröße von $> 95\%$ und haben infolgedessen einen annähernd maximalen Wert.

Die so definierte Sigmoidfunktion ist für alle Größen gleich, ohne eine nötige Vorgabe einer individuellen physikalischen Parametrierung. Mithilfe dieser können die signaltheoretischen Eigenschaften (Standardabweichung) aller gemessenen Größen zu einer einzigen Kenngröße $J_S(\text{LMR}j)$ im Stellbereich $\text{LMR}j$ zusammengeführt werden:

$$J_S(\text{LMR}j) = \frac{J_S(\sigma_{x_1}^*) + \dots + J_S(\sigma_{x_e}^*) + J_S(\sigma_{y_1}^*) + \dots + J_S(\sigma_{y_a}^*) + J_S(\sigma_{l_i}^*)}{E + A + 1} \quad (4.39)$$

Die Kenngröße $J_S(\text{LMR}j)$ entspricht dabei einer über alle Größen und alle Abtastwerte gemittelten Standardabweichung der Messwerte im relevanten Stellbereich $\text{LMR}j$.

Gesamtkriterium zur iterativen Stellbereichsanpassung

Ausgehend von den für jeden lokalen Stellbereich LMR j erfassten Kenngrößen J_Q und J_S wird ein Gesamtkriterium J_A bestimmt, indem beide Kriterien linear überlagert werden:

$$J_A(\text{LMR } j) = \frac{J_Q(\text{LMR } j) + J_S(\text{LMR } j)}{2} \quad (4.40)$$

Der Zusammenhang der drei Kriterien ergibt sich zu:

$$\left. \begin{array}{l} J_Q(\text{LMR } j) \rightarrow 0 \text{ und } J_S(\text{LMR } j) \rightarrow 0 \\ J_Q(\text{LMR } j) \rightarrow 1 \text{ und } J_S(\text{LMR } j) \rightarrow 0 \\ J_Q(\text{LMR } j) \rightarrow 0 \text{ und } J_S(\text{LMR } j) \rightarrow 1 \\ J_Q(\text{LMR } j) \rightarrow 1 \text{ und } J_S(\text{LMR } j) \rightarrow 1 \end{array} \right\} \Rightarrow \begin{array}{l} J_A(\text{LMR } j) \rightarrow 0 \\ 0 < J_A(\text{LMR } j) < 1 \\ J_A(\text{LMR } j) \rightarrow 1 \end{array} \quad (4.41)$$

Der Stellbereich wird bei der (stationären) Vermessung nacheinander entsprechend der Anzahl an Iterationen geteilt. In jedem Schritt wird dabei das sich ergebende Kriterium J_A in allen j Teilbereichen bestimmt. Die nächste Teilung wird dann in dem Teilbereich mit dem kleinsten Kriterium vorgenommen. Ein kleines Kriterium J_A entspricht dabei einer geringen Modellgüte und einer geringen Standardabweichung bzw. einem geringen Rauschen der relevanten gemessenen Größen. Nachdem alle Teilungen durchgeführt worden sind, werden in dem Aktorstellbereich mit dem kleinsten Kriterium J_A neue Messpunkte generiert, welche mit modellunabhängigen oder modellbasierten Verfahren festgelegt werden. Diese werden mit der nächsten Iteration vermessen.

Mithilfe dieser Adaption werden neue Messpunkte im Betriebsbereich mit der schlechtesten Modellgüte vermessen. Dabei wird die Varianz der Messgrößen in der Adaption mitberücksichtigt. Der wesentliche Vorteil ist, dass die automatisierte Adaptionsmethodik den Fokus nicht nur in qualitativ schlechte Bereiche legt, sondern auch die Variation der zu modellierenden Ausgangsgrößen einbezieht. Infolgedessen erfolgt das automatisierte Vermessen seltener in Bereichen, welche z.B. aufgrund der hohen Variation der Zylinderdrücke p_{Zyl} nah an der Zündgrenze liegen, als in normalen Betriebsbereichen.

4.3.3 Modellunabhängige Bestimmung der Messpunkte

Eine Möglichkeit zur Erstellung von modellunabhängigen Messpunkten ist die Verwendung von klassischen Messplänen (siehe Abschnitt 4.1.3). Nachteilig ist dabei, dass bereits gemessene Messpunkte nicht einkalkuliert werden, und es demzufolge zu vielen ungewollten Wiederholungspunkten kommen kann.

Besser geeignet sind iterativ raumfüllende Messpläne. Dabei werden in dem reduzierten Stellbereich, ausgehend von den bereits gemessenen Messpunkten, neue Messpunkte generiert. Die einfachste Möglichkeit ist, eine Sobol-Sequenz vollständig für die Anzahl an maximal zu vermessenden Punkten aufzustellen. In jeder Iteration wird ergo die entsprechende Anzahl an Punkten der Sequenz für die Messung übernommen [94]. Eine Alternative kann aus einer Anzahl an zufälligen Kandidaten alle Euklidischen- (bei gleich skalierten Achsen) bzw. Mahalanobis-Entfernungen (bei verschieden skalierten Achsen) zu allen bereits gemessenen Messpunkten

bestimmt werden. Aus diesen Abständen wird für jeden Kandidatenpunkt der nächste bereits existierende Nachbarpunkt bestimmt. Der Punkt, mit der größten Entfernung zum Nachbarpunkt ist der gesuchte raumfüllende Punkt [58]. Nachteilig ist hierbei, dass nicht der tatsächlich entfernteste Punkt bestimmt wird, sondern nur der entfernteste aus dem zufälligen Kandidatensatz. Weiterhin ist das vorgeschlagene Vorgehen nicht iterativ raumfüllend, da in jedem Schritt immer nur einer der nächsten Nachbarpunkte aller Messpunkte als zu vermessender Punkt hinzugenommen wird. Eine weitere Möglichkeit besteht darin, die Messpunkte direkt und nicht basierend auf einem Kandidatensatz zu bestimmen, indem der euklidische Abstand zwischen den alten und neu hinzukommenden Punkten mithilfe eines genetischen Algorithmus optimiert wird [105, 80].

Eine schnelle Bestimmung des Messkandidaten, welcher tatsächlich am entferntesten von allen bereits gemessenen Punkten liegt, erfolgt mithilfe des DTS-Verfahrens (*delaunay triangulation sampling*). Dieses basiert auf der in Abschnitt 3.5.1 beschriebenen konkaven Triangulation. Mit Hilfe der von dem reduzierten Aktorstellbereich aufgespannten Eckpunkten sowie den innerhalb dessen liegenden bereits vermessenen Punkten wird die konkave Triangulation bestimmt. Die Hinzunahme der Eckpunkte ermöglicht dabei die vollständige Abdeckung des reduzierten Aktorstellbereichs durch die konkave Triangulation. Insofern als bei der Triangulation jedes Simplex keine weitere Stützstelle von Nachbarsimplexen enthält, kann der den Raum iterativ füllende Punkt bestimmt werden. Hierfür wird zuerst das Simplex mit dem größten Volumen bestimmt. Dieses entspricht somit genau dem größten freien Raum im reduzierten Stellbereich. Der gesuchte raumfüllende Messpunkt M stimmt mit dem geometrischen Schwerpunkt c des volumenmäßig größten Simplex, aufgespannt durch $P + 1$ Punkte v_i überein:

$$M = c = \frac{1}{P + 1} \sum_{i=1}^{P+1} v_i \quad (4.42)$$

Die Bestimmung des Schwerpunkts des größten Simplex entspricht dabei nicht, wie im abstandsmaximierenden Verfahren der Maximierung des Abstands zu bereits vermessenen Punkten (= Mittelpunkt), sondern der mittigen Platzierung eines Punktes im größten noch nicht vermessenen Raum. Wird die Schwerpunktbestimmung iterativ für die gewünschte Anzahl an Messplanpunkten wiederholt, so wird der Raum iterativ mit neuen raumfüllenden Messplanpunkten abgedeckt. Die Triangulation muss dabei, trotz des neuen Punktes, nicht wiederholt werden. Vielmehr kann die geometrische Eigenschaft ausgenutzt werden, dass die durch Hinzufügen des Schwerpunktes entstehenden $P + 1$ inneren Simplexe gleiche Volumen haben. Somit kann das größte Simplex in der Triangulation durch $P + 1$ kleinere mit jeweils gleichen Volumen ersetzt werden, ohne dass die gesamte Triangulation neu berechnet werden muss. Das Vorgehen zur Bestimmung eines neuen raumfüllenden Messpunktes ist zusammenfassend in Bild 4.12 für das DTS- sowie das in [58] beschriebene abstandsmaximierende Verfahren dargestellt.

Bild 4.13 stellt die Messpläne für zwei Stellgrößen ($P = 2$) nach der 1., 3. und 6. Iteration für die zwei iterativ raumfüllenden Verfahren (abstandsmaximierendes Verfahren und DTS-Verfahren) dar. Zusätzlich sind die jeweiligen Histogramme der zugehörigen Triangulationen

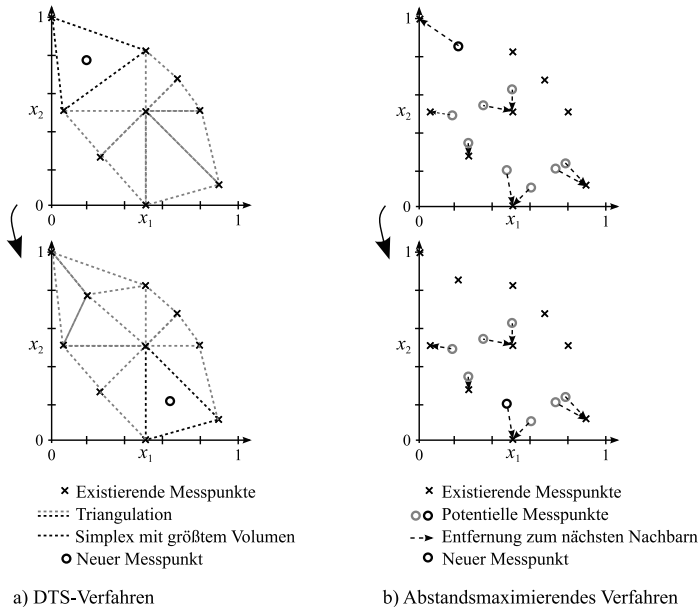


Bild 4.12: Vorgehen zur iterativen Raumauffüllung vom a) DTS-Verfahren und b) abstandsmaximierenden Verfahren.

dargestellt. Ausgehend vom initialen Versuchsplan mit 9 Messpunkten werden in jeder Iteration 5 neu zu vermessende Messpunkte innerhalb des konkaven Raumes hinzugenommen. Beide Verfahren generieren neue Messpunkte unter Einbezug der bereits vorhandenen. Es wird deutlich, dass mit dem DTS-Verfahren eine bessere iterative Raumauffüllung des konkaven Raumes erreicht wird. Dies ist bereits zu Beginn einer Vermessung, wenn nur wenige Messpunkte vorhanden sind, der Fall. Die Histogramme der auftretenden Dreiecksflächen veranschaulichen dies. Eine raumfüllend optimale Verteilung wäre im Histogramm durch einen einzigen Balken maximaler Höhe gekennzeichnet. Die Histogramme des DTS-Verfahrens weisen in allen Iterationen sowohl eine kleinere Streubreite als auch eine höhere Anzahl der am häufigsten auftretenden Dreiecksflächen auf. Vorteilhaft ist nicht nur die bessere Raumauffüllung, sondern auch die Unabhängigkeit von einem Kandidatensatz, der beim abstandsmaximierenden Verfahren erst noch geeignet gewählt werden muss.

4.3.4 Modellbasierte Bestimmung der Messpunkte

Ist die Modellstruktur des vermessenen Prozesses bekannt, kann durch iterative modellabhängige Verfahren eine zielführende Festlegung der Messpunkte erfolgen. Diese generieren im Gegensatz zu den modellunabhängigen Verfahren neue Messpunkte, indem nicht nur alle bereits gemessenen Punkte, sondern auch das aktuelle Modell berücksichtigt werden. Dabei wird das

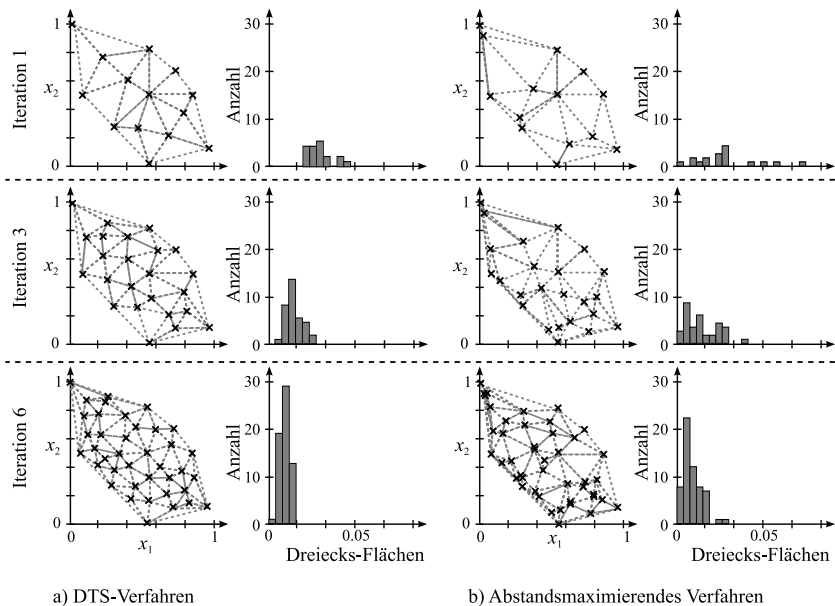


Bild 4.13: Vergleich der iterativen Raumauffüllung in verschiedenen Iterationen anhand der Messpunkte und des Histogramms der Triangulationsflächen vom a) DTS-Verfahren und b) abstandsmaximierenden Verfahren.

Modell für ein Raster an neuen Messpunktkandidaten \mathbf{K} im reduzierten Stellbereich ausgewertet und je nach Verfahren neue Messpunkte abgeleitet. Die neu zu vermessenden Messplanpunkte ergeben sich somit direkt aus den aktuell vorhandenen Prozessinformationen. Im Rahmen der hier vorgestellten Online-Methodik kommen vier verschiedene Verfahren der iterativen modellbasierten Messplanerstellung zum Einsatz, welche im Folgenden näher beschrieben werden.

Messpläne basierend auf der Varianz eines Modellkomitees

Eine Methode zur Messplanerweiterung ist das Query-by-Committee (Komitee-Fragestellung, [123]). Bei der Komitee-Fragestellung werden, basierend auf dem Vergleich der für die Messpunktkandidaten \mathbf{K} simulierten Modellausgänge von verschiedenen Modellen mit unterschiedlichen Strukturen (Modellkomitees) neue Messpunkte generiert. Das Modellkomitee kann dabei aus Polynommodellen erhöhter Ordnung, Netzmodellen erhöhter Teilung sowie Modellen alternativer Art (z.B. Support-Vektor-Modellen oder Gauß'schen Prozessmodellen) sowie einer beliebigen Kombination aller bestehen. Die Zusammenstellung des Komitees hängt dabei stark von dem zu vermessenden Prozess ab. Je unbekannter bzw. stärker nichtlinear der Prozess ist, desto „breiter“ sollte das Komitee aufgestellt werden, und verschiedene Arten der Modellierung sollten zum Einsatz kommen. Ist der Prozess nur leicht nichtlinear bzw. die Eignung eines

z.B. Polynomansatzes zur Modellierung annehmbar, sollte aus Gründen der Rechenzeit auf den Einsatz von alternativen Modellierungsarten verzichtet werden. Basierend auf den verwendeten Modellen wird für jeden neuen Stellgrößenkandidaten $\mathbf{u}_{N+1} \in \mathbf{K}$ ein Modellausgang \hat{y}_{QbC} des Modellkomitees, welcher der Mittlung der Modellausgänge \hat{y}_j aller J verwendeten Modelle entspricht, bestimmt:

$$\hat{y}_{\text{QbC}} = \frac{1}{J} \sum_{j=1}^J \hat{y}_j \quad (4.43)$$

Durch die Modellierung der Messdaten mit verschiedenen Modellstrukturen ergeben sich unterschiedliche Modellausgänge für die simulierten Kandidaten $\mathbf{u}_{N+1} \in \mathbf{K}$. Die Idee des Modellkomitees besteht darin, dass der Stellgrößenkandidat \mathbf{u}_{N+1} , bei dem die resultierende Varianz der Modellausgänge am größten ist, also für den gilt

$$\mathbf{M} = \arg \max_{\mathbf{u}_{N+1} \in \mathbf{K}} \left(\frac{1}{J-1} \sum_{j=1}^J (\hat{y}_j - \hat{y}_{\text{QbC}})^2 \right) \quad (4.44)$$

als neu zu vermessender Punkt \mathbf{M} mit in das Design übernommen wird. Zu beachten ist dabei, dass die Varianz aus Modellen unterschiedlicher Struktur bestimmt wird, welche zwar alle die aktuellen Messdaten, aber nicht zwangsläufig alle den Prozess darstellen. Andererseits ist aber im Rahmen der Online-Vermessung, speziell in den ersten Iterationen, die tatsächlich vorliegende Prozessstruktur aufgrund der wenigen Messdaten noch nicht genau bestimmt, wodurch der Einsatz eines Modellkomitees zur Bestimmung von neuen Messpunkten speziell in diesem Bereich von Vorteil ist. Generell eignet sich ein Modellkomitee zur Simulation von unbekannten (nicht bereits vermessenen) Modelleingängen bei noch nicht endgültig festgelegter Modellstruktur, da hierbei die positiven Eigenschaften des Komitees überwiegen.

Messpläne basierend auf der erwarteten Änderung der Modellparameter

Die Messplanerweiterung, basierend auf der erwarteten Änderung der Modellparameter (MPC, model parameter change), erstellt neue Messpunkte anhand des iterativen Modells. Dabei wird der zur Identifizierung verwendete Datensatz um den Stellgrößenkandidaten $\mathbf{u}_{N+1} \in \mathbf{K}$ erweitert, indem angenommen wird, dass der unbekannte (aber zur Schätzung benötigte) gemessene Prozessausgang für den neuen Datenpunkt \mathbf{u}_{N+1} dem jeweils simulierten Komiteeausgang, und somit einem fiktiv gemessenen Wert bei der jeweiligen Stellgrößenkombination \mathbf{u}_{N+1} entspricht:

$$\begin{aligned} \mathbf{u}_+ &= \begin{bmatrix} \mathbf{u} \\ \mathbf{u}_{N+1} \end{bmatrix}; \mathbf{y}_+ = \begin{bmatrix} \mathbf{y} \\ \hat{y}_{\text{QbC}} \end{bmatrix} \\ \Rightarrow \Psi_+ &= \begin{bmatrix} \Psi \\ \Psi_{N+1} \end{bmatrix} = \begin{bmatrix} 1 & u_{1,1} & u_{2,1} & \cdots & u_{R-1,1} \\ 1 & u_{1,2} & u_{2,2} & \cdots & u_{R-1,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & u_{1,N} & u_{2,N} & \cdots & u_{R-1,N} \\ 1 & u_{1,N+1} & u_{2,N+1} & \cdots & u_{R-1,N+1} \end{bmatrix} \end{aligned} \quad (4.45)$$

Mit diesem um den Stellgrößenkandidat $\mathbf{u}_{N+1} \in \mathbf{K}$ erweiterten Datensatz zur Modellbildung wird eine neue Parameterschätzung unter Beibehaltung der Modellstruktur durchgeführt:

$$\hat{\boldsymbol{\theta}}_+ = (\boldsymbol{\Psi}_+^T \boldsymbol{\Psi}_+)^{-1} \boldsymbol{\Psi}_+^T \mathbf{y}_+ \quad (4.46)$$

Der Stellgrößenkandidat, bei dem die Änderung der Parameter im Vergleich zu den ursprünglichen Modellparametern am größten ist, für den demnach gilt

$$\mathbf{M} = \arg \max_{\mathbf{u}_{N+1} \in \mathbf{K}} (\|\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_+\|) \quad (4.47)$$

wird als neu zu vermessende Stellgröße in den Messplan übernommen. Hierdurch wird der neue Messpunkt für das in dieser Iteration tatsächlich identifizierte Modell gezielt in Bereiche positioniert. Eine große Änderung der Parameter deutet dabei auf ein nicht robustes Modell hin. Es ist von einer noch unpassenden Modellstruktur oder einer unpassenden Messpunktverteilung auszugehen.

Messpläne basierend auf der erwarteten Änderung der Modellgüte

Auch die Methode, welche Messpunkte basierend auf der erwarteten Änderung der Modellgüte definiert, nutzt den Modellkomiteeausgang als fiktiven Messwert zur Erweiterung der Datensätze sowie die in der aktuellen Iteration verwendete Modellstruktur zur Schätzung der Modellparameter. Dabei wird zwischen zwei Verfahren unterschieden. Bei dem MQI-Verfahren (model quality increase) ergibt sich der Messpunktkandidat \mathbf{M} aus den Stellgrößenkandidaten $\mathbf{u}_{N+1} \in \mathbf{K}$ mit dem größten sich ergebenden Fehler zwischen den beiden Modellen. Hierfür wird ausgehend von den neuen Modellparametern der neue Prozessausgang für jeden Stellgrößenkandidaten \mathbf{u}_{N+1} simuliert:

$$\hat{\mathbf{y}}_+(\mathbf{u}_{N+1}) = \boldsymbol{\Psi}_+(\mathbf{u}_{N+1}) \hat{\boldsymbol{\theta}}_+(\mathbf{u}_{N+1}) \quad (4.48)$$

Der neue Messpunktkandidat \mathbf{M} folgt aus der Maximierung der Fehlerdifferenz J_{RMSE} zwischen den beiden Modellen zu

$$\mathbf{M} = \arg \max_{\mathbf{u}_o \in \mathbf{K}} (J_{\text{RMSE}} - J_{+, \text{RMSE}}) \quad (4.49)$$

$$= \arg \max_{\mathbf{u}_o \in \mathbf{K}} \left(\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} - \sqrt{\frac{1}{N+1} \sum_{i=1}^{N+1} (y_{+,i} - \hat{y}_{+,i})^2} \right) \quad (4.50)$$

Ein abnehmender Fehler bzw. eine zunehmende Güte deutet auf eine bereits passende Modellstruktur, jedoch fehlende Messdaten im Stellraum hin, was bedeutet, dass der die Güte maximierende Messpunkt gewünscht ist.

Aus dem gleichen Kriterium kann mit dem zweiten Verfahren MQD (model quality decrease) ein weiterer zu vermessender Messpunktkandidat \mathbf{M} definiert werden, indem die Minimierung der Fehlerdifferenz zwischen den beiden Modellen erfolgt:

$$\mathbf{M} = \arg \min_{\mathbf{u}_{N+1} \in \mathbf{K}} (J_{\text{RMSE}} - J_{+, \text{RMSE}}) \quad (4.51)$$

Ein maximal zunehmender Fehler bzw. eine maximal abnehmende Güte deutet, vice versa zur maximal zunehmenden Güte, auf eine nicht passende Struktur des verwendeten Modells hin. Auch diese Messpunkte sind im Rahmen der iterativen Vermessung von Bedeutung und werden der Vermessung hinzugefügt, um eine Anpassung der Modellbildungsstruktur bei der folgenden Iteration vornehmen zu können.

Messpläne basierend auf den Modelloptima

Die Messplananpassung, basierend auf lokalen Sollwertkennfeldoptima (MO, model optima), erweitert den Messplan um das Optimum (Minimum) einer Zielfunktion in einem Motorarbeitspunkt. Ziel der Optimierung ist, die iterativ Identifizierten Modelle zu nutzen um optimale Stellgrößenkombinationen zur Steuerung zu finden. Bei diesen ist der Kraftstoffverbrauch minimiert bzw. das Motordrehmoment maximiert unter Einhaltung der gesetzlichen Grenzwerte für die Emissionen.

Bei der stationären lokalen Vermessung wird das im konstanten Arbeitspunkt ($n_{\text{Mot},i}, q_{\text{HE},j}$) der Sollwertkennfelder bestimmte lokale Stellgrößenoptimum als Messpunkt in den Messplan aufgenommen. Dabei ist zu berücksichtigen, dass in dem jeweiligen Arbeitspunkt der Kraftstoffverbrauch $\dot{m}_{f,o}$ konstant ist, womit in der Zielfunktion eine Minimierung des spezifischen Kraftstoffverbrauchs erfolgt. Aus den Anforderungen in Abschnitt 4.1.5 folgt die zu minimierende Zielfunktion für den spezifischen Kraftstoffverbrauch in $\left[\frac{\text{g}}{\text{kWh}}\right]$ in jedem Arbeitspunkt $o = 1, \dots, O$ zu:

$$\mathbf{M} = \mathbf{u}_{\text{opt,lokal},o} = \arg \min_{\mathbf{u}_o \in \mathcal{V}R_o} \left(J_{\text{lokal}}(\mathbf{u}_o) \right) \quad (4.52)$$

$$\text{mit } J_{\text{lokal}}(\mathbf{u}_o) = \gamma \frac{3600 \left[\frac{\text{s}}{\text{h}}\right] \left(w_1 \dot{m}_{f,o} \left[\frac{\text{g}}{\text{s}}\right] + w_2 \dot{m}_{\text{NO}_x,o}(\mathbf{u}_o) \left[\frac{\text{g}}{\text{s}}\right] + w_3 \dot{m}_{\text{RuB},o}(\mathbf{u}_o) \left[\frac{\text{g}}{\text{s}}\right] \right)}{P_o(\mathbf{u}_o) [\text{kW}]}$$

Diese Optimierung entspricht einer Minimierung der NO_x - und Ruß-Emissionen und Maximierung der Motorleistung

$$P_o(\mathbf{u}_o) [\text{W}] = M_{\text{Mot},o}(\mathbf{u}_o) [\text{Nm}] \cdot n_{\text{Mot},o} \left[\frac{1}{\text{s}} \right] \cdot 2\pi \quad (4.53)$$

in jedem Arbeitspunkt o bei konstantem Kraftstoffverbrauch

$$\dot{m}_{f,o} \left[\frac{\text{g}}{\text{s}} \right] = \frac{4}{2} \left[\frac{\text{Zyl}}{\text{U}} \right] \cdot q_{\text{HE},o} \left[\frac{\text{mm}^3}{\text{Zyl}} \right] \cdot \rho_{\text{Diesel}} \left[\frac{\text{g}}{\text{mm}^3} \right] \cdot n_{\text{Mot},o} \left[\frac{\text{U}}{\text{s}} \right]. \quad (4.54)$$

Die Gewichte w_j können dabei entweder basierend auf Erfahrungswerten gemäß

$$w_{1,\dots,3} \in [0, 1] \text{ und } \sum_{j=1}^3 w_j = 1 \quad (4.55)$$

vorgegeben werden oder zyklusspezifisch, indem das Verhältnis der Zyklusgrenzwerte berücksichtigt wird ($180 \text{ mg/km NO}_x : 5 \text{ mg/km RuB} = 36 : 1$ für EURO 5):

$$w_1 = \frac{1}{3}, w_2 = \frac{2}{3} \cdot \frac{1}{37}, w_3 = \frac{2}{3} \cdot \frac{36}{37}. \quad (4.56)$$

Durch die einfache Beschreibung der konkaven oder konvexen Variationsräume mithilfe der entsprechenden Triangulation gemäß Abschnitt 3.5.1, können beide im Rahmen der Optimierung einbezogen werden. Hierfür wird die Lage der optimierten Größen im Bezug zum Variationsraum ausgewertet:

$$\gamma = \begin{cases} 1 & \text{für } \mathbf{u}_o \in \text{VR}_o \\ 100 & \text{für } \mathbf{u}_o \notin \text{VR}_o \end{cases} \quad (4.57)$$

Durch den Faktor γ wird eine Optimierung der Stellgrößen in Bereichen außerhalb des zulässigen Variationsraums vermieden.

Die Bestimmung des Minimums der Zielfunktion erfolgt mit einem lokalen Optimierer (z.B. Simplex-Algorithmus Nelder-Mead [118]) ausgehend von verschiedenen Startpunkten für die definierten Gewichte. Diese Startpunkte sind sowohl das vorherige Optimum der letzten Iteration als auch die Schwerpunkte der geteilten lokalen Aktorstellbereiche. Durch die immer kleiner werdenden Bereiche und das immer genauer bestimmte Optimum wird das globale lokale Optimum immer besser bestimmt, obwohl nur ein lokaler Optimierer zum Einsatz kommt.

Soll die Optimierung nicht für eine definierte Gewichtung der Zielgrößen erfolgen, kann alternativ mithilfe eines Mehrzieloptimierers (z.B. Genetischer-Algorithmus NSGA-II [27]) das Optimum der Zielfunktion bestimmt werden. Aufgrund des Zielkonfliktes zwischen den NO_x - und Ruß-Emissionen, d.h. geringere NO_x -Emissionen werden durch höhere Ruß-Emissionen erreicht, kann keine eindeutige Lösung der Optimierung bestimmt werden. Vielmehr wird ein Pareto-Optimum bestimmt, welches den Lösungen der Optimierung für jeweils verschiedene Gewichte entspricht.

Die stationär globale Vermessung, bei der mehrere Arbeitspunkte des Motors vermessen werden, ermöglicht, das globale Optimum der Sollwertkennfelder in der Messplanerweiterung einzubeziehen. Dazu werden die optimierten Stellgrößenkombinationen aller Arbeitspunkte in die jeweils zugehörigen Messpläne aufgenommen. Die resultierende zu minimierende Zielfunktion für die globale Optimierung mit $o = 1, \dots, O$ Arbeitspunkten lautet

$$\begin{aligned} \mathbf{M} &= \mathbf{u}_{\text{opt,global}} = \arg \min_{\mathbf{u} \in V_R} \left(J_{\text{global}}(\mathbf{u}) \right) \\ \text{mit } J_{\text{global}}(\mathbf{u}) \left[\frac{\text{g}}{\text{kWh}} \right] &= b_c(\mathbf{u}) \left[\frac{\text{g}}{\text{kWh}} \right] \cdot (\gamma + \text{mean}(|\Gamma \mathbf{u}|)) \\ &= \frac{3600 \left[\frac{\text{s}}{\text{h}} \right] \cdot \left(\sum_{o=1}^O w_o [\text{s}] \cdot \dot{m}_{f,o}(\mathbf{u}_o) \left[\frac{\text{g}}{\text{s}} \right] \right)}{\sum_{o=1}^O w_o [\text{s}] \cdot P_o(\mathbf{u}_o) [\text{kW}]} \cdot (\gamma + \text{mean}(|\Gamma \mathbf{u}|)) \end{aligned} \quad (4.58)$$

Die Zielfunktion beinhaltet neben dem spezifischen Kraftstoffverbrauch $b_c(\mathbf{u})$ zusätzlich eine Gewichtung in Abhängigkeit von der Lage der optimierten Größen im Variationsraum (γ , siehe (4.57)) sowie in Abhängigkeit von der Glattheit des resultierenden Kennfeldes ($\text{mean}(|\Gamma \mathbf{u}|)$, siehe Abschnitt H). Eine nachträgliche Glättung der Optima der Sollwertkennfelder würde dazu führen, dass die dazugehörigen optimierten Stellgrößen $\mathbf{u}_{\text{opt,global}}$ verschoben werden, und diese dadurch nicht mehr ihren Optima entsprechen.

Zusätzlich müssen die sich aus den gesetzlichen Zyklusgrenzwerten ergebenden Beschränkungen gemäß

$$\frac{\sum_{o=1}^O w_o[s] \cdot \dot{m}_{\text{NO}_x,o}(\mathbf{u}_o) \left[\frac{\text{g}}{\text{s}} \right]}{s_{\text{Zyklus}} [\text{km}]} \leq \dot{m}_{\text{NO}_x,\text{lim}} \left[\frac{\text{g}}{\text{km}} \right] \quad (4.59)$$

$$\frac{\sum_{o=1}^O w_o[s] \cdot \dot{m}_{\text{Ruß},o}(\mathbf{u}_o) \left[\frac{\text{g}}{\text{s}} \right]}{s_{\text{Zyklus}} [\text{km}]} \leq \dot{m}_{\text{Ruß},\text{lim}} \left[\frac{\text{g}}{\text{km}} \right] \quad (4.60)$$

bei der Optimierung einbezogen werden. Dabei wird jedem Arbeitspunkt o ein eigenes Gewicht w_o in Sekunden zugeordnet, welches sich aus der Verweildauer in dem jeweiligen Zyklusarbeitspunkt ergibt.

Auch hier kommt aus Gründen der schnelleren Bestimmung des Optimums ein lokaler Optimierer zum Einsatz, welcher ausgehend von verschiedenen Startpunkten das jeweilige Optimum bestimmt. Die Startpunkte entsprechen dabei allen lokalen Optima der ersten Iteration für jeweils jeden Arbeitspunkt. Ab der $P + 1$ Vermessungsiteration (entspricht der zweiten Vermessung des ersten Arbeitspunktes) werden die globale optimierten Stellgrößen als Startwerte vorgegeben.

Durch die Hinzunahme der aus den Optimierungen bestimmten iterativen lokalen und globalen Optima in den Messplan wird erreicht, dass die jeweiligen Optima vermessen werden und sowohl eine iterativ immer genauere Bestimmung als auch eine direkte Validierung der Optima am Prüfstand erfolgt.

4.3.5 Zielgerichtete Messplanerweiterung im Rahmen der Online-Methodik

Durch die geeignete Kombination der iterativen modellbasierten Verfahren kann eine zielgerichtete Vermessung im Rahmen der Online-Methodik erfolgen. Dabei wird einbezogen, dass die Vermessung mit wenigen Messpunkten startet und unter Anpassung des identifizierten Modells wiederholt wird. Die Modellstruktur ist dabei vor allem zu Beginn der Vermessung noch einer starken Änderung unterworfen, welche im weiteren Verlauf geringer wird. Um dieses Verhalten im Rahmen der iterativen Vermessung zu berücksichtigen, erfolgt die Bestimmung des Messplans nach definierten Kriterien. In der ersten Iteration ($n = 1$) erfolgt die Erstellung des Messplans mit einem Hüllraster, um den Variationsraum zu bestimmen. In der zweiten Iteration ($n = 2$) erfolgt die Vermessung basierend auf dem raumfüllenden DTS-Verfahren, da zu dem Zeitpunkt der Messplanerweiterung noch keine Modellinformation vorlag. Durch die Vermessung von modellunabhängigen Messpunkten wird der Fokus der neuen Daten auf die Anpassung der Modellstruktur gelegt. Die weitere Vermessung ab Iteration $n > 2$ erfolgt gemäß Bild 4.14.

Solange die Modellstruktur sich stark ändert (d.h. solange entweder die statische Polynomordnung S sich erhöht oder die Netzmodellteilung erweitert wird, wodurch die Anzahl M an lokalen Teilmodellen zunimmt) erfolgt die Messplanerweiterung basierend auf dem DTS- und MQD-Verfahren. Durch das DTS-Verfahren werden raumfüllende Messpunkte im gesamten Stellbereich platziert, um mehr Prozessinformation aus diesen zu erhalten. Zusätzlich wird

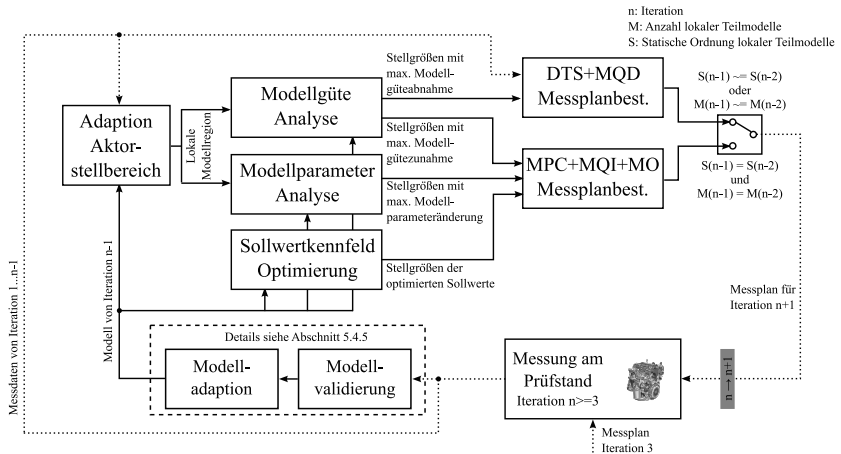


Bild 4.14: Darstellung der einzelnen Schritte der zielgerichteten Messplanerweiterung für eine Iteration ≥ 3 . Die Auswahl der Verfahren zur Bestimmung der Messpunkte erfolgt basierend auf der Erkennung einer sich ändernden Modellstruktur (Anzahl lokaler Teilmodelle M sowie lokale Polynomordnung S).

mit dem MQD-Verfahren im schlechtesten reduzierten Stellbereich die Messpunkte hinzugefügt, welche auf eine noch nicht optimale Struktur deuten. Ändert sich die Modellstruktur nicht mehr signifikant, erfolgt die Messplanerweiterung mit Messpunkten, welche sich aus den MPC-, MQI-, und MO-Verfahren ergeben. Durch die Berücksichtigung der maximalen Modellparameteränderung und Zunahme der Modellgüte wird der Fokus auf eine gute Messpunktverteilung für die vorliegende Modellstruktur gelegt. Zusätzlich gibt es durch den Einbezug der Modelloptima eine verbesserte Modellgüte in Bereichen der gesuchten Optima.

4.3.6 Beispiel der Online-Versuchsplanung

In diesem Abschnitt werden für einen simulierten Prozess die Online-Messplanerweiterung mit einem klassischen raumfüllenden Verfahren zur Erstellung von Messplänen verglichen. Der zu bestimmende Prozess wird dabei durch die statische Peaks-Funktion simuliert, welche einer Überlagerung von mehreren zweidimensionalen Gaußglocken entspricht (siehe Bild 4.15). Durch den Einsatz dieser Beispielfunktion kann die Güte der zielgerichteten Messplanerweiterung anhand einer stark nichtlinearen Funktion bewertet werden. Ein Beispiel für den Verbrennungsmotor folgt in Kapitel 6.

Als Modell kommt ein LOLIMOT-Netzmodell zum Einsatz, wobei die Netzmodellteilung auf 30 beschränkt wird und für beide Messplanerweiterungen gleich angenommen wird. Durch den Vergleich der Strategien zur Versuchsplanung anhand von linearen Netzmodellen mit gleichen Teilungen kann der Einfluss der Messplanerweiterung unabhängig von dem zugrunde geleg-

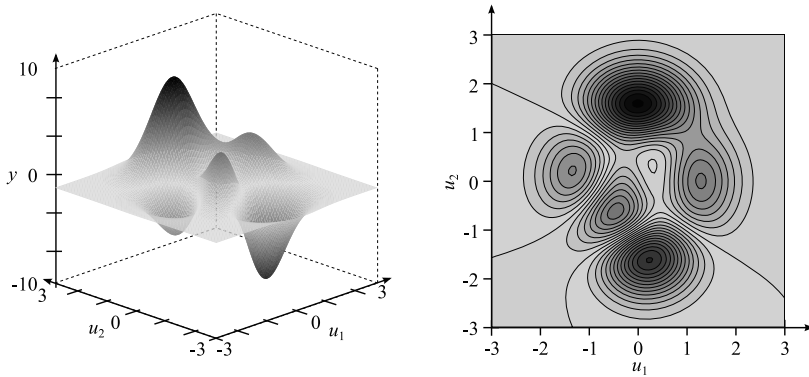


Bild 4.15: Darstellung der Ausgangsgröße sowie der resultierenden Höhenlinien für den nichtlinearen statischer Peaks-Beispielprozess mit zwei Eingangs- und einer Ausgangsgröße zum Vergleich der Verfahren zur Erstellung von Messplänen.

ten Modell bewertet werden. Dabei werden iterativ beispielhaft 3 neue Messpunkte hinzugenommen, beim klassischen Verfahren gemäß dem raumfüllenden Sobol und im zielgerichteten Online-Verfahren gemäß dem kombinierten DTS- und MQD-Verfahren. Die Netzmodellstruktur mit maximal 30 Teilungen wird in den jeweiligen Iterationen manuell definiert. Hierdurch werden keine neuen Messpunkte nach den MO-, MPC- und MQI-Verfahren generiert, jedoch ermöglicht dies den Vergleich der Verfahren zur Messpunktauswahl ohne dass eine eventuell unterschiedliche Modellbildung einen Einfluss auf das Ergebnis hat. In jeder Iteration wird zusätzlich mit einem weiteren Messdatensatz, bestehend aus 2809 LHS- und 2809 Raster-Messpunkten, das jeweilige Modell validiert.

Die sich nach der letzten Iteration ergebenden Modellausgangsgrößen für die Vermessung mit einem Sobol-Messplan und dem zielgerichteten Online-Messplan sind in Bild 4.16 a) als *measured vs. predicted* Plot dargestellt. Weiterhin ist in b) der Verlauf der Trainings- und Validierungsgüte über die Iterationen aufgetragen. Sowohl in der Darstellung der Modellausgangsgröße als auch im Verlauf der Trainings- und Validierungsgüte wird deutlich, dass die Messplanerweiterung im Rahmen der Online-Methodik zielgerichteter verläuft als das klassische Vorgehen. Die verbleibenden Modellfehler im Training und in der Validierung sind mit dem Online-Verfahren geringer. Der Unterschied zwischen den beiden Verfahren ist dabei nur gering, folgt jedoch ausschließlich aufgrund der besseren Platzierung der Messpunkte im Versuchsraum, da die zugrundeliegenden Modellstrukturen bei beiden Verfahren dieselben sind. Durch die schnellere Konvergenz gegen die maximal erreichbare Modellgüte kann bei der Online-Methodik bereits mit weniger Messpunkten ein besseres Modell bestimmt werden.

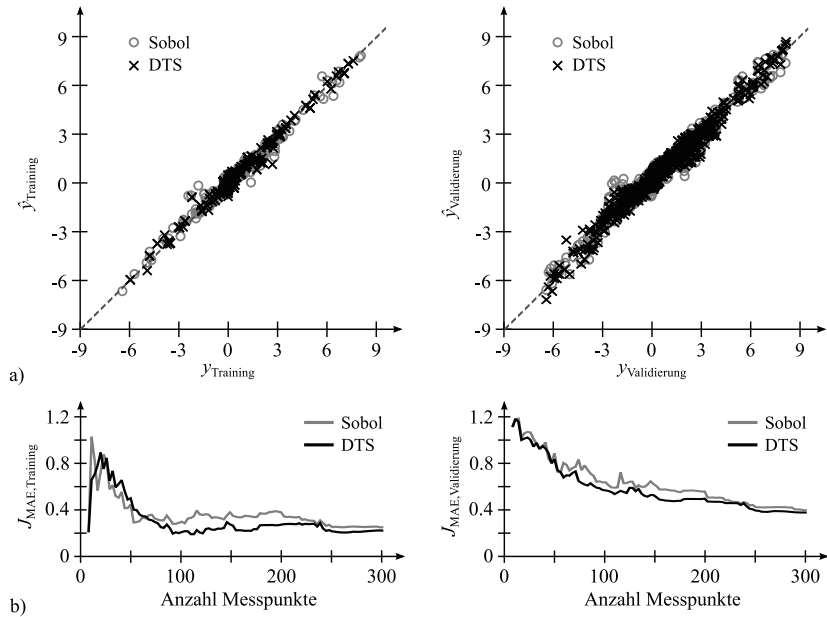


Bild 4.16: Vergleich der Vermessung mit einem Sobol- sowie DTS-Messplan für einen simulierten statischen Peaks-Prozess. Die identifizierten Modelle aus beiden Messplänen weisen eine hohe Güte auf (a). Bei Betrachtung der Güte in Abhängigkeit von der Anzahl an Messpunkten wird das zielgerichtete Vorgehen des DTS-Messplans deutlich, bei welchem der Modellfehler kleiner ist und schneller in Richtung des mit dieser Modellstruktur minimal möglichen Modellfehlers konvergiert (b).

4.4 Zusammenfassung

Der Entwurf einer geeigneten Versuchsplanung für einen Prozess sowie die eingesetzte Modellbildung stellt einen wesentlichen Aspekt der Vermessung dar. Die in diesem Kapitel entwickelten Methoden zur Prozessanregung mit Kurzzeitsignalen (siehe Abschnitt 4.2) sowie zur gezielten rekursiven Messplanerweiterung (siehe Abschnitt 4.3) ermöglichen eine zielgerichtete Versuchsplanung und automatische Anpassung an den zugrundeliegenden Prozess, so dass nur wenig Information vorab verfügbar sein müssen.

Die Anregung des dynamischen Prozesses erfolgt mit dem beschriebenen DBS-Signal. Hierdurch wird nicht nur das statische, sondern auch das dynamische Prozessverhalten zielgerichtet angeregt unter flexibler prozessbezogener Parametrierung des Signals. Somit können höhere Modellqualitäten bzw. kürzere Signallängen erreicht werden, wodurch Messzeit eingespart werden kann.

Für eine zielgerichtete Messplanerweiterung wurde ein kombinierter aktiv lernender Messplan entworfen, welcher eine iterative Anpassung des Aktorstellbereichs mit einer iterativen Messplanerweiterung innerhalb des Variationsraums kombiniert. Die Online-Versuchsplanung berücksichtigt, dass die Vermessung mit wenigen Messpunkten startet und mit einer iterativen Anpassung des identifizierten Modells wiederholt wird.

Die Stellbereichsanpassung bestimmt den in der nächsten Iteration zu vermessenden Stellbereich basierend auf einem Gesamtkriterium, welches die globale Modellgüte mit der Messgrößenvarianz kombiniert. Hierdurch wird der Fokus der Versuchsplanung automatisiert auf Bereiche mit einer geringen Varianz der gemessenen Größen und geringen Güte des modellierten Prozesses gerichtet, da dort der größte Informationsgehalt für die Identifikation gewonnen werden kann.

Die Messplanerweiterung erfolgt sowohl raumfüllend, um noch nicht genügend modellierte Bereiche zu entdecken, als auch modellbasiert, um Messpunkte zielgerichtet an den Stellen des größten Informationsgewinns zu platzieren. Die entwickelten Methoden zur Messplanerweiterung sind dabei zielgerichtet auf das aktuell bestimmte Modell entworfen. Das DTS-Verfahren nutzt die Triangulation des konkaven Variationsraumes und erreicht bei geringem Rechenaufwand eine gleichmäßigere Raumfüllung des konkaven Stellbereichs als klassische raumfüllende Verfahren. Das DTS-Verfahren wird mit den entwickelten modellbasierten MQD, MPC, MQI und MO-Verfahren kombiniert, um eine optimierte Modellstruktur, eine gute Messpunktverteilung für die aktuelle Modellstruktur sowie eine hohe Modellgüte im Bereich der Stellgrößen-optima zu erreichen.

5 Online-Identifikation von mathematischen Verbrennungsmotor-Modellen

In diesem Kapitel wird eine Methode zur Erweiterung der klassischen Modellbildung von Prozessen um die Berücksichtigung des richtungsabhängigen Verhaltens vorgestellt. Durch die Ergänzung der Regressormatrix zur Parameterschätzung um eine richtungsabhängige Schaltfunktion wird die automatische Identifikation der Richtung von statischen oder dynamischen Prozessen (z.B. Prozesse mit Hysteresen oder Prozesse mit verschiedenen Zeitkonstanten je nach Stellrichtung) ermöglicht. Weiterhin werden die für den Einsatz der experimentellen Modellbildung im Rahmen der Online-Vermessung notwendigen Erweiterungen hin zu einer Online-Identifikation vorgestellt. Die Besonderheiten bei der Analyse und Bewertung von iterativen Online-Modellen, das heißt die Erhöhung der Komplexität bei gleichzeitiger Steigerung der Messpunktezahl, werden dargestellt. Zur Modellbildung werden um die Onlinefähigkeit erweiterte lokal polynomiale Netzmodelle (ONLOPOMOT) eingesetzt, welche für den Sonderfall von einem einzigen Teilmodell einem Polynommodell (ONPOLY) entsprechen. Neben einer iterativen Adaption der lokalen Modellparameter wird dabei auch die iterative Anpassung der Modellstruktur, bestehend aus den Regressoren sowie der Netzmodellteilung, berücksichtigt. Die Regressorauswahl erfolgt kontinuierlich durch die Kombination der LASSO-Parameterschätzung mit einer Erhöhung der maximal zulässigen Regressorordnung bei Bedarf. Die Bestimmung der passenden Regressoren wird durch die Teilungsanpassung des Netzmodells ergänzt, wobei dies neben der klassischen Vergrößerung bei Bedarf auch eine Verkleinerung des Netzmodells enthalten kann.

Mathematische Modelle von Verbrennungsmotoren kommen heute aufgrund der gestiegenen Komplexität in den verschiedenen Bereichen der Entwicklung wie z.B. Systementwurf, Regelungsentwurf, Basis-Kalibrierung oder Fehlererkennung zum Einsatz. Zusätzlich können diese durch den Einsatz von Modellen als virtuelle Sensoren für verschiedene Ausgangsgrößen wie z.B. das Motormoment oder die Abgastemperatur genutzt werden[19]. Hierdurch werden schwer messbare Größen erfasst sowie teure reale Sensoren ersetzt, sodass die Anzahl an verbauten realen Sensoren kleiner und somit die Kosten gering gehalten werden.

Bei der Vermessung zur Bestimmung von Verbrennungsmotormodellen kommen zwei Arten der mathematischen Modellbildung zum Einsatz: Die theoretische und die experimentelle Modellbildung. Bei der theoretischen Modellbildung werden die inneren physikalischen Zusammenhänge des betrachteten Systems modelliert und beschrieben. Dabei wird mithilfe von vereinfachenden Annahmen ein Gleichungssystem, bestehend aus Bilanzgleichungen, Schaltungleichungen, konstitutiven und/oder phänomenologischen Gleichungen, aufgestellt, dessen Lösung das Prozessverhalten analytisch beschreibt. Zur Bestimmung dieser Modelle werden keine Messdaten benötigt. Nachteil der so entstehenden White-Box-Modelle ist, dass sie bereits für

einfache Prozesse nichtlinear und aufwendig zu bestimmen sowie meist für den Einsatz in Echtzeitsystemen aufgrund der langen Simulationsdauer ungeeignet sind.

Die rein experimentellen Modelle, welche bei der Online-Methodik überwiegend eingesetzt werden, beschreiben den Zustand eines Systems anhand von Ein- und Ausgangssignalen. Da die inneren physikalischen Zusammenhänge dieser Black-Box-Modelle unbekannt sind, muss die Prozessstruktur entweder bekannt sein oder geeignet angenommen werden. Dem Nachteil der umfangreichen Versuchsplanung und -durchführung steht der Vorteil der guten Eignung zum Einsatz in Echtzeitsystemen sowie geringem benötigten Vorwissen gegenüber, siehe [62, 63, 67].

Die Mischform beider Modellbildungsarten, die semiphysikalische Modellbildung (Grey-Box-Modelle), kombiniert die Vorteile beider Verfahren, indem, basierend auf vereinfachenden Annahmen, der Prozess mit seinen physikalischen Zusammenhängen beschrieben und mithilfe von Messdaten bestimmt wird. Dabei werden, da die physikalische Struktur nur teilweise bekannt ist, die dominanten Prozesseigenschaften physikalisch modelliert und die Modellparameter entsprechend mit experimentellen Modellen angepasst, um eine hohe Güte der Modelle zu erreichen. Der Einsatz von semiphysikalischen Modellen ist auch im Rahmen der Online-Methodik möglich, indem im Voraus die Prozesseigenschaften physikalisch modelliert werden, und das entstehende Vorwissen als Modellstruktur vorgegeben wird.

Im Folgenden wird die experimentelle Modellbildung verfolgt, da das Ziel dieser Arbeit darin liegt, eine Methodik mit möglichst wenig Vorwissen zu entwickeln. Die Modellbildung erfolgt im Anschluß an die Messdatenerfassung im Rahmen einer Vermessung (siehe Bild 5.1). Eine Vielzahl von Methoden sind in letzten Jahren für die experimentelle Modellbildung entstanden (siehe Abschnitt 5.1). Für den Einsatz im Rahmen einer Online-Vermessung ist zu berücksichtigen, dass die Datensatzgröße zunimmt und sich dadurch das Bias-Varianz-Dilemma auflöst (siehe Abschnitt 5.2). Die daraus folgenden Anpassungen der Modellanalyse werden in Abschnitt 5.3 beschrieben. Ist aufgrund der geringen Validierungsgüte eine Anpassung des Modells nötig, kann durch eine Parameterschätzung mit integrierter Regressoradaption sowie rekursiver Anpassung von Netzmodellstrukturen (siehe Abschnitt 5.4) ein Modell identifiziert werden, welches mit zunehmender Datensatzgröße einen sinkenden Bias- und Varianzfehler hat.

5.1 Grundlagen der experimentellen Modellbildung

Die experimentelle Modellbildung (Identifikation) bestimmt das mathematische Modell, bestehend aus Modellstruktur und Modellparametern, anhand der gemessenen Ein- und Ausgangssignale. Dabei bleiben physikalische Zusammenhänge unberücksichtigt und die Prozessstruktur muss, falls benötigt, entweder bekannt sein (z.B. aus Vorwissen vorangegangener Vermessungen) oder geeignet gewählt werden. Im Rahmen der Vermessung von Verbrennungsmotoren kommen verschiedene Arten der parametrischen und nicht-parametrischen Modelle zum Einsatz. Zu den parametrischen Modellen zählen neben einfachen Kennfeldern auch polynomiale Modelle, neuronale Netzmodelle oder Gauß'sche Prozessmodelle. Bei parametrischen Pro-

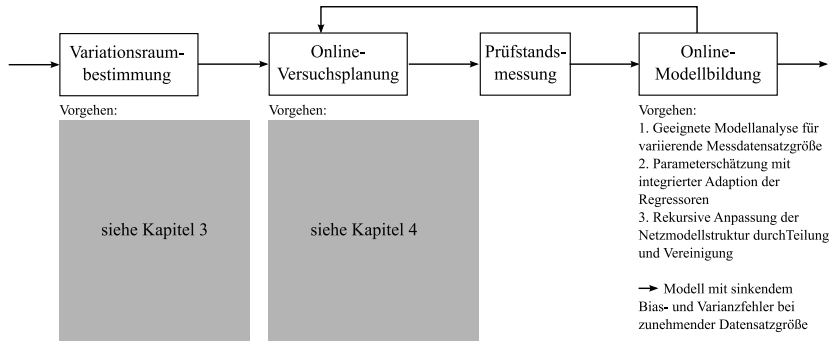


Bild 5.1: Vereinfachte Darstellung der einzelnen Schritte einer Vermessung und daraus resultierend das Vorgehen zum Entwurf einer durchgängigen Online-Methode im Rahmen dieser Arbeit bzw. in diesem Kapitel.

zessmodellen werden die benötigten Parameter mithilfe der gemessenen Daten und geeigneter Methoden so bestimmt, dass der Ausgangs- oder der Gleichungsfehler minimal wird. Zu den nicht-parametrischen Modellen zählen Kennfeldmodelle, welche die dem Prozess zugrundeliegenden Parameter nur implizit in Form einer Wertematrix enthalten. Weiterhin wird zwischen stationären und dynamischen Modellen unterschieden.

5.1.1 Parameterschätzung

Die Parameter von Modellen, welche linear in ihren Parametern sind, und für diese und für den zu modellierenden Prozess gilt

$$\begin{aligned} y &= \Psi \Theta + n \\ \hat{y} &= \Psi \hat{\Theta} \end{aligned} \quad (5.1)$$

können mithilfe der Methode der kleinsten Quadrate (engl. least-squares) geschätzt werden. Hierfür wird die modellstruktur- und messpunktabhängige Regressormatrix

$$\Psi = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \cdots & x_{R-1,1} \\ 1 & x_{1,2} & x_{2,2} & \cdots & x_{R-1,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,N} & x_{2,N} & \cdots & x_{R-1,N} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & x_1 & x_2 & \cdots & x_{R-1} \end{bmatrix} \quad (5.2)$$

aufgestellt. Ein Regressor ist dabei eine statistisch unabhängigen Variable, welche einen zu bestimmenden Einfluss auf die zu erklärende Größe y ausübt. Im Fall von Prozessen, deren Ausgangsgröße linear von den Stellgrößen abhängt, gilt für die Regressoren x_o (mit $o = 1, \dots, R-1$)

$$x_o \in \{u_1, \dots, u_P\} \quad (5.3)$$

während bei Prozessen mit einem nichtlinearen Zusammenhang zwischen Stell- und Ausgangsgröße zusätzlich Kombinationen sowie Potenzen der Stellgrößen als Regressoren vorkommen

können:

$$\mathbf{x}_o \in \{\mathbf{u}_1, \dots, \mathbf{u}_P, \mathbf{u}_1^2, \dots, \mathbf{u}_P^2, \mathbf{u}_1 \mathbf{u}_2, \dots, \mathbf{u}_1 \mathbf{u}_P\}. \quad (5.4)$$

Zusammen mit dem Ausgangsgrößenvektor $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ wird ausgehend von dem Residuum oder Gleichungsfehler e_i (Prädiktionsfehler-Methode [100]) mit $i = 1, \dots, N$

$$e_i = y_i - \hat{y}_i = y_i - \Psi_i \hat{\boldsymbol{\theta}} \quad (5.5)$$

die zu minimierende LS-Zielfunktion nach [41] aufgestellt:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg \min_{\hat{\boldsymbol{\theta}}} \left(\mathbf{e}^T(\mathbf{x}, \hat{\boldsymbol{\theta}}) \mathbf{e}(\mathbf{x}, \hat{\boldsymbol{\theta}}) \right) = \arg \min_{\hat{\boldsymbol{\theta}}} \left(\sum_{i=1}^N e_i^2(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) \right) \\ &= \arg \min_{\hat{\boldsymbol{\theta}}} \left(\sum_{i=1}^N \left(y_i(\mathbf{x}_i) - \hat{y}_i(\mathbf{x}_i, \hat{\boldsymbol{\theta}}) \right)^2 \right) \\ &= \arg \min_{\hat{\boldsymbol{\theta}}} \left(\sum_{i=1}^N \left(y_i - \Psi_i \hat{\boldsymbol{\theta}} \right)^2 \right) \end{aligned} \quad (5.6)$$

Dabei wird die Summe der Fehlerquadrate zwischen Prozess und Modell minimiert. Die Lösung dieses Optimierungsproblems kann geschlossen erfolgen, siehe z.B. [70, 120], indem das LS-Gleichungssystem

$$\hat{\boldsymbol{\theta}} = (\Psi^T \Psi)^{-1} \Psi^T \mathbf{y} \quad (5.7)$$

gelöst wird. Hierbei existiert eine eindeutige Lösung nur bei $N \geq R$, folglich wenn mehr Messpunkte als zu bestimmende Parameter vorhanden sind, da ansonsten ein unterbestimmtes Gleichungssystem entsteht.

Das direkte LS-Verfahren zur Parameterschätzung kann nur zur im quadratischen Mittel konsistenten Parameterschätzung von Modellen angewandt werden, welche linear in Ihren Parametern sind, deren Ausgangsgrößen entweder ungestört ($\mathbf{n} = 0$) oder nur mit stationären unkorrelierten Störsignalen ($E\{\mathbf{n}\} = 0$) überlagert sind, und deren Eingangsgrößen nicht-fehlerbehaftet sind und den Prozess genügend anregen. Ansonsten liefert die LS-Schätzung keine Bias-freie Lösung, siehe z.B. [62, 70].

Je nach vorliegenden Voraussetzungen der Messungen bzw. Modelle sowie zur Lösung des Bias-Problems existieren verschiedene Erweiterungen der LS-Schätzung bzw. alternative Verfahren. Diese sind zusammenfassend in Anhang G sowie in [70] aufgeführt.

Im dem Fall, dass Messdaten rekursiv aufgezeichnet werden und somit zu Beginn der Parameterschätzung nicht vollständig vorhanden sind, kann das *recursive* LS-Verfahren (RLS, [70]) eingesetzt werden. Dabei werden die Modellparameter $\hat{\boldsymbol{\theta}}(k+1)$ nicht basierend auf allen Messdaten neu bestimmt, sondern es werden die alten Modellparameter $\hat{\boldsymbol{\theta}}(k)$ mit lediglich den neusten Messdaten aktualisiert:

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \boldsymbol{\gamma}(k) \left(y(k+1) - \Psi(k+1) \hat{\boldsymbol{\theta}}(k) \right) \quad (5.8)$$

Auf diese Weise kann der Rechenaufwand zur Bestimmung der Parameter, vor allem bei großen Datenmengen welche rekursiv erfasst werden, reduziert werden, und eine Online-Identifikation wird ermöglicht.

5.1.2 Modellbildungsverfahren für stationäre Prozesse

Alle im Folgenden beschriebenen Verfahren zur Identifikation von Modellen bestimmen bei der Parameterschätzung mithilfe einer Zielfunktion die im Mittel besten Parameter. Dieses Vorgehen ist prinzipiell für alle Prozesse geeignet, bei denen keine Richtungsabhängigkeit vorhanden ist. Eine Methode zur Erweiterung der Modellbildung um die Berücksichtigung der Richtungsabhängigkeit ist in Anhang I beschrieben.

Polynomiale Modelle

Polynome sind für stationäre Prozesse mathematische Modelle mit der einfachsten Modellstruktur und basieren auf der Taylor-Reihenentwicklung der nichtlinearen Funktion y . Entsprechend einer Taylor-Reihe [22] wird eine genauere Approximation durch eine erhöhte Ordnung S des Polynoms erreicht, die jedoch eine gestiegene Komplexität zur Folge hat. Für ein Polynommodell der statischen Ordnung S sowie mit P Stellgrößen gilt die allgemeine Gleichung:

$$\hat{y}_i = b_0 + \sum_{o=1}^{R-1} b_j x_{o,i} = b_0 + \sum_{o=1}^{R-1} b_j u_{1,i}^k u_{2,i}^l \dots u_{P-1,i}^m u_{P,i}^n \quad (5.9)$$

mit $(k + l + \dots + m + n) \leq S$

Dabei ergeben sich die Regressoren $x_{o,i}$ durch Kombination von Stellgrößen $u_{1,i}, \dots, u_{P,i}$ mit beliebigen Potenzen, solange die Summe aller Potenzen innerhalb eines Regressors kleiner als die statische Ordnung ist. Für eine statische Ordnung von $S = 2$ und $P = 3$ Stellgrößen folgt exemplarisch

$$\begin{aligned} \hat{y}_i = & b_0 + b_1 u_{1,i}^1 u_{2,i}^0 u_{3,i}^0 + b_2 u_{1,i}^0 u_{2,i}^1 u_{3,i}^0 + b_3 u_{1,i}^0 u_{2,i}^0 u_{3,i}^1 + \\ & b_4 u_{1,i}^1 u_{2,i}^1 u_{3,i}^0 + b_5 u_{1,i}^0 u_{2,i}^1 u_{3,i}^1 + b_6 u_{1,i}^1 u_{2,i}^0 u_{3,i}^1 + \\ & b_7 u_{1,i}^2 u_{2,i}^0 u_{3,i}^0 + b_8 u_{1,i}^0 u_{2,i}^2 u_{3,i}^0 + b_9 u_{1,i}^0 u_{2,i}^0 u_{3,i}^2 \end{aligned} \quad (5.10)$$

Zu beachten ist, dass z.B. der Regressor $u_{1,i}^1 u_{2,i}^1 u_{3,i}^1$ nicht Teil der Gleichung ist, da die Summe seiner statischen Ordnung 3 und somit $> S$ ist.

Die Anzahl aller möglichen Regressoren, welche den Modelleingangsgrößen entsprechen, kann mithilfe einer Analogie zur Kombinatorik bestimmt werden. Nach Feller [39] folgt die Anzahl an k -Tupeln, welche aus nicht-negativen ganzen Zahlen mit der Summe $\leq n$ bestehen (unter der Annahme das n und k positive ganze Zahlen sind), aus dem folgenden Binomialkoeffizienten:

$$\begin{aligned} \binom{n+k-1}{k-1} &= \frac{n+k-1}{1} \cdot \frac{n+k-2}{2} \cdot \frac{n+k-3}{3} \cdot \dots \cdot \frac{n+k-(k-1)}{k-1} \\ &= \frac{(n+k-1)(n+k-2)(n+k-3) \dots (n+1)}{(k-1) \cdot \dots \cdot 3 \cdot 2 \cdot 1} \\ &= \frac{\prod_{j=1}^{k-1} (n+j)}{\prod_{j=1}^{k-1} j} = \prod_{j=1}^{k-1} \frac{n+j}{j} \end{aligned} \quad (5.11)$$

Mit der Analogie, dass ein k -Tupel bei dem Polynomialmodell gemäß Gleichung (5.9) aus $P + 1$ -Größen (alle P Stellgrößen sowie der Offset) besteht und n der Summe S aller Potenzen innerhalb eines Tupels entspricht, folgt die Anzahl R an Regressoren in Abhängigkeit von der Anzahl n an Stellgrößen und der statischen Ordnung zu

$$R = \binom{S+P}{P} = \prod_{j=1}^P \frac{(S+j)}{j}. \quad (5.12)$$

Die unbekannten Modellparameter $\hat{\Theta} = [b_0, b_1, \dots, b_{R-1}]^T$ können bei Polynomen mithilfe der Methode der kleinsten Quadrate bestimmt werden, da sie linear in ihren Parametern sind. Hierfür wird die LS-Schätzgleichung nach Gleichung (5.7) aufgestellt und gelöst. Die Lösung des Gleichungssystems entspricht den gesuchten Polynomparametern $\hat{\Theta}$. Mit den Parametern kann das Polynommodell gemäß Gleichung (5.9) aufgestellt und der Modellausgang für beliebige Eingangsgrößen bestimmt werden.

Neuronale Netzmodelle

Neuronale Netzmodelle sind universelle Funktionsapproximatoren, welche im Bezug auf das Signalverhalten an die biologischen Neuronen des menschlichen Gehirns angelehnt sind (siehe [120, 129]). Zum Einsatz kommen dabei verschiedene Arten der neuronalen Netze wie z.B. Multilayer-Perzeptron-Netze (MLP) oder Radial-Basis-Funktions-Netze (RBF). Diese haben den Nachteil, dass sie nicht physikalisch interpretierbar sind und zumeist eine geringe Konvergenzgeschwindigkeit bzw. eine große Rechenzeit benötigen.

Netzmodelle mit schnell konvergierenden Parameterschätzmethoden und einer geringen Rechenzeit sind LOLIMOT (Local Linear Model Tree, [119]), LOPOMOT (Local Polynomial Model Tree, [146]) und HHT (Hinging Hyperplane Tree, [161]). Die spezielle Netzstruktur erlaubt eine automatische Adaption der Modellstruktur an das nichtlineare statische und dynamische Prozessverhalten, indem lokal eine lineare oder polynomiale Funktionsapproximation erfolgt.

LOLIMOT-Netzmodelle stellen einen nichtlinearen Prozess durch eine Überlagerung mehrerer lokal linearer (bzw. falls ein Offset vorhanden ist lokal affiner) Teilmodelle (LLM) dar. Der Aufteilung des Eingangsraums in j -Teilgebiete mit den jeweiligen Längen Δj der Hyperrechteckseiten in denen die einzelnen Teilmodelle gültig sind (Struktur des LOLIMOT-Netzes), wird dabei über eine automatische achsen-orthogonale Teilung erreicht (siehe Bild 5.2 a)). Der Zusammenhang zwischen den $j = 1, \dots, M$ lokal linearen Teilmodellen

$$\begin{aligned} \hat{y}_j(x_i) &= b_{0,j} + b_{1,j}x_{1,i} + \dots + b_{R-1,j}x_{R-1,i} = \Psi_i \hat{\Theta}_j \\ \text{mit } x_{o,i} &\in \{u_{1,i}, \dots, u_{P,i}\} \\ \text{und } i &= 1, \dots, N \text{ sowie } o = 1, \dots, R-1 \end{aligned} \quad (5.13)$$

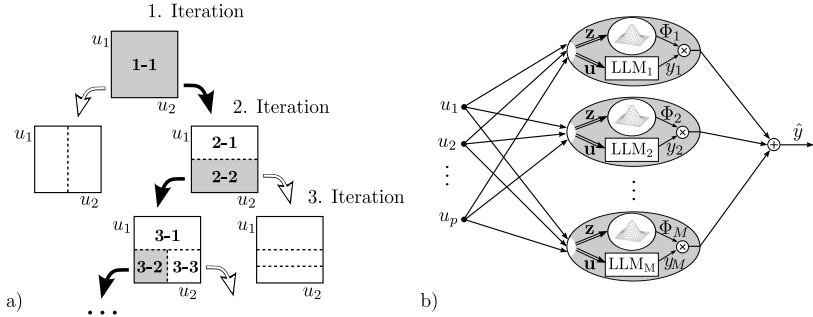


Bild 5.2: a) Achsen-orthogonaler Teilungsalgorithmus zur Strukturbestimmung sowie b) die resultierende Struktur eines LOLIMOT Modells. Die lokalen Modelle werden mit einer Gauß'schen Gewichtungsfunktion zu einer globalen Modellausgangsgröße überlagert. Die Bereiche der lokalen Teilmodelle werden dabei mit dem Teilungsalgorithmus bestimmt [146].

wird für die jeweiligen lokalen Regressoren \mathbf{x}_i und lokalen Modellparameter $\hat{\boldsymbol{\theta}}_j$ über normierte Gaußfunktionen Φ_j als Gewichts- bzw. Gültigkeitsfunktionen hergestellt:

$$\Phi_j(\mathbf{z}_i) = \frac{\exp\left(-\frac{1}{2} \left\| \frac{\mathbf{z}_i - \mathbf{z}_{0,j}}{\boldsymbol{\sigma}_j^2} \right\|_2^2\right)}{\sum_{j=1}^M \exp\left(-\frac{1}{2} \left\| \frac{\mathbf{z}_i - \mathbf{z}_{0,j}}{\boldsymbol{\sigma}_j^2} \right\|_2^2\right)} \quad \text{mit } \mathbf{z}_i = \mathbf{x}_i. \quad (5.14)$$

Hierbei ist $\mathbf{z}_{0,j}$ die Mittelpunktskoordinate des j -ten Teilmodells und $\boldsymbol{\sigma}_j^2$ die Varianz der j -ten Gaußfunktionen. Das nichtlineare Gesamtmodell ergibt sich gemäß Bild 5.2 b) als Überlagerung aller lokal linearen Teilmodelle mit der entsprechenden Gewichtungsfunktion aus

$$\hat{y}_{\text{LOLIMOT}}(\mathbf{x}_i) = \sum_{j=1}^M \Phi_j(\mathbf{z}_i) \hat{y}_j(\mathbf{x}_i). \quad (5.15)$$

Hierbei stellen die von \mathbf{z} abhängigen Gaußfunktionen den nichtlinearen und die von \mathbf{x} abhängigen lokalen Modelle den linearen Zusammenhang des Prozesses dar. Mithilfe der Standardabweichung der Gaußfunktionen $\boldsymbol{\sigma}_j$ kann dabei die Überlagerung der einzelnen Teilmodelle variiert werden. Je kleiner die Standardabweichung gewählt wird, desto schärfer ist die Abgrenzung der einzelnen Teilmodelle. Eine zu kleine Standardabweichung führt dazu, dass das Gesamtmodell \hat{y}_{LOLIMOT} sich fast ausschließlich aus dem aktuellen lokalen Modell \hat{y}_j ergibt. Durch die geringen Übergangsbereiche wird der Gesamtmodellausgang unstetig. Wird die Standardabweichung zu groß gewählt, haben alle Teilmodelle den gleichen Einfluss auf den Gesamtmodellausgang \hat{y}_{LOLIMOT} und dieser entspricht demzufolge dem Mittelwert aller lokal linearen Teilmodelle. Als Richtwert kann für die Standardabweichung $\boldsymbol{\sigma}_j = \frac{\Delta \mathbf{j}}{3}$ angenommen werden [120]. Hierbei steht $\Delta \mathbf{j}$ für die Seitenlängen in alle Dimensionen des j -ten Teilmodells.

Die Parameter $\hat{\boldsymbol{\theta}}_j$ der j lokalen Teilmodelle folgen mit dem WLS-Verfahren zu (z.B. [120, 70])

$$\hat{\boldsymbol{\theta}}_j = (\boldsymbol{\Psi}^T \boldsymbol{W}_j \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}^T \boldsymbol{W}_j \boldsymbol{y}. \quad (5.16)$$

Die Regressormatrix $\boldsymbol{\Psi}$ und der Ausgangsgrößenvektor \boldsymbol{y} folgen aus Gleichung (5.2). Die Gewichtungsmatrix \boldsymbol{W}_j den j -ten lokalen Modells entspricht dabei einer Diagonalmatrix mit den Gewichtungsfunktionen $\Phi_j(\boldsymbol{z}_i)$:

$$\boldsymbol{W}_j = \begin{bmatrix} \Phi_j(\boldsymbol{z}_1) & 0 & \cdots & 0 \\ 0 & \Phi_j(\boldsymbol{z}_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Phi_j(\boldsymbol{z}_N) \end{bmatrix}. \quad (5.17)$$

LOPOMOT erweitert die Modellbildung mit lokal linearen Netzmodellen zu lokal polynomialen Netzmodellen. Dabei werden die lokalen Modelle \hat{y}_j anstelle von Gaußfunktionen mit einer pyramidalen Gewichtungsfunktion $\Phi_j(\boldsymbol{z})$ zu dem globalen Modellausgang \hat{y}_{LOPOMOT} überlagert (siehe Bild 5.3):

$$\begin{aligned} \hat{y}_{\text{LOPOMOT}}(\boldsymbol{x}_i) &= \sum_{j=1}^M \Phi_j(\boldsymbol{z}_i) \hat{y}_j(\boldsymbol{x}_i) \\ \text{mit } \hat{y}_j(\boldsymbol{x}_i) &= b_{0,j} + b_{1,j}x_{1,i} + \dots + b_{R-1,j}x_{R-1,i} = \boldsymbol{\Psi}_i \hat{\boldsymbol{\theta}}_j \\ \text{und } x_{o,i} &\in \{u_{1,i}, \dots, u_{P,i}, u_{1,i}^2, u_{1,i}u_{2,i}, \dots, u_{P,i}^S\} \\ \text{und } i &= 1, \dots, N \text{ sowie } o = 1, \dots, R-1. \end{aligned} \quad (5.18)$$

LOPOMOT stellt somit, im Gegensatz zu LOLIMOT, den nichtlinearen Zusammenhang des Prozesses mit den lokalen Teilmodellen und den linearen Zusammenhang mit den Gewichtungsfunktionen dar.

Die Anzahl R an signifikanten aus den insgesamt Q möglichen Regressoren der lokalen Polynome wird durch die Kombination einer schrittweisen Auswahlstrategie mit der lokalen Mallows' C_p -Statistik bestimmt. Die Anzahl M an lokalen Teilmodellen wird zusätzlich mit einer globalen C_p -Statistik bewertet. Durch die Kombination der positiven Eigenschaften von Netz- und Polynommodellen wird eine Reduktion des Modellfehlers erreicht. Dabei wird die Komplexität des Netzmodells verringert (Anzahl an Modellteilungen), indem die lokale Komplexität erhöht wird (Anzahl an lokalen Modellregressoren).

Ein weiterer Ansatz, welcher Polynom- und Netzmodelle kombiniert, ist POLYMOT [8]. Hierfür wird iterativ die Erhöhung der lokalen Polynomordnung mit der Erhöhung der Netzmodellteilung verglichen und die Komplexitätserhöhung, welche eine bessere Modellgüte zur Folge hat, durchgeführt. Die Regressorauswahl erfolgt dabei mit heuristischen Tuningparametern.

Hinging hyperplane trees (HHT,[21, 161]) erweitern die Modellbildung mit LOLIMOT, indem nicht eine achsenorthogonale Teilung zur Strukturbestimmung durchgeführt wird, sondern die optimale Richtung (bzw. der optimale Winkel) der Teilung mithilfe einer nichtlinearen Optimierung bestimmt wird. Neuere Ansätze nutzen lokale Modellnetzwerke mit inkrementierenden Netzmodellstrukturen, welche Ihre Komplexität mit zunehmenden Daten erhöhen [57].

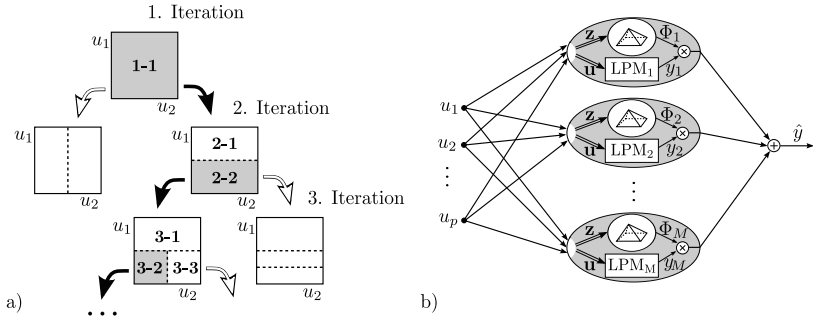


Bild 5.3: Die Netzstruktur eines LOPOMOT wird durch einen Teilungsalgorithmus bestimmt (a). Der globale Modellausgang folgt durch die gewichtete Überlagerung von adaptiven lokalen polynomialen Modellen (b) [146].

Es gibt verschiedenen Erweiterungen der hier vorgestellten neuronalen Netzmodelle, z.B. Verfahren mit einer Clustering anstelle der achsenorthogonalen Teilung [121], Verfahren mit angepassten Parameterschätz- und Teilungsalgorithmen [74, 73] oder Verfahren mit Erweiterungen für dynamische Modellierungen [173], auf die im Folgenden jedoch nicht weiter eingegangen wird.

Nichtlineare Modellbildungsverfahren

Mit der immer größeren verfügbaren Rechenleistung werden zunehmend auch Modellbildungsverfahren angewandt, welche nichtlinear in ihren Modellparametern sind.

Multilayer Perceptron Netzmodelle (MLP,[16]) sind weiterentwickelte neuronale Netzmodelle, welche überwiegend bei der Mustererkennung aber auch im Rahmen der Motorapplikation eingesetzt [108, 154] werden. MLP-Netzmodelle sind universelle Approximatoren, welche aus einer Eingangs-, einer versteckten sowie einer Ausgangsschicht bestehen und eine tanh-Funktion als Aktivierungsfunktion nutzen. Für den Modellausgang folgt:

$$\hat{y} = \Theta_A f_{\tanh}(\Theta_E \Psi) \quad (5.19)$$

Die Modellparameter Θ_E und Θ_A werden dabei mithilfe von nichtlinearen Optimierern bestimmt, wobei die Wahl eines geeigneten Optimierungsalgorithmus stark abhängig von dem zugrundeliegenden Prozess ist [16].

Eine weitere, aufgrund der steigenden Rechen- und Speicherleistung von Motorsteuergeräten, zunehmend eingesetzte Modellierungsart sind Kernel-Modelle, zu denen Gauß'sche- und Support-Vektor-Modelle gehören. Gauß'sche-Modelle [101, 126] sind nicht-parametrische, statistische Modelle, bei denen sich der Modellausgang aus überlagerten Gaußglocken ergibt, welche jedem Messpunkt zugeordnet werden. Die Modellbildung erfolgt dabei unter Annahme eines normalverteilten Rauschens

$$y = \hat{y}(x) + e \text{ mit } e \sim N(0, \sigma_n^2) \quad (5.20)$$

durch die Maximierung des Logarithmus der marginalen Likelihood-Funktion $p(y|\hat{\Theta})$

$$\hat{\Theta} = \arg \max_{\hat{\Theta}} \left(\ln p(y|\hat{\Theta}) \right) \quad (5.21)$$

welche dem Integral der Likelihood-Funktion entspricht [126, 14]:

$$\begin{aligned} \ln p(y|\hat{\Theta}) &= \ln \left(\int p(y|\hat{y}, \hat{\Theta}) p(\hat{y}|\hat{\Theta}) d\hat{y} \right) \\ &= -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{K} + \sigma_n^2 \mathbf{I}| - \frac{1}{2} \mathbf{y}^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} \end{aligned} \quad (5.22)$$

Als Kovarianzfunktion \mathbf{K} wird dabei in der Regel eine mehrdimensionale Gaußglocke verwendet:

$$\begin{aligned} \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) &= \sigma_f^2 \exp \left(-\frac{1}{2} (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M} (\mathbf{x}_i - \mathbf{x}_j) \right) \\ &\text{mit } \mathbf{M} = \text{diag} (l_1^{-2}, l_2^{-2}, \dots, l_R^{-2}) \end{aligned} \quad (5.23)$$

Die beim Training des Modells optimierten Parameter $\hat{\Theta}$ sind die charakteristischen Längenmaße l_1, \dots, l_R sowie die Ausgangsgrößenvarianz σ_f^2 bei gegebener Rauschvarianz σ_n^2 . Die charakteristischen Längenmaße sind dabei ein Maß für die Entfernung in der jeweiligen Dimension, ab welcher der Modellausgangswert sich signifikant ändern darf [126]. Kleine Längenmaße führen zu einem welligen Modellausgang und umgekehrt. Die Optimierung erfolgt dabei basierend auf den Trainingsdaten mit der Quasi-Newton-Methode, indem die Ableitungen der Funktion (5.22) nach dem jeweiligen Parameter gebildet werden. Der simulierte Prozessausgang für einen neuen Datenpunkt \mathbf{x}_o ergibt sich als Mittelwert und Kovarianz des Prozesses zu:

$$\begin{aligned} \bar{\hat{y}}_+ &= \mathbf{K}(\mathbf{x}_o, \mathbf{x}) [\mathbf{K}(\mathbf{x}, \mathbf{x}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{y} \\ \text{cov} \hat{y}_+ &= \mathbf{K}(\mathbf{x}_o, \mathbf{x}_o) - \mathbf{K}(\mathbf{x}_o, \mathbf{x}) [\mathbf{K}(\mathbf{x}, \mathbf{x}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{K}(\mathbf{x}, \mathbf{x}_o) \end{aligned} \quad (5.24)$$

Gauß'sche Modelle stellen universelle Approximatoren dar, wodurch sie zur Modellbildung generell geeignet sind, jedoch ist der Zeitbedarf zur Modellbildung und Simulation durch die benötigte Optimierung der Parameter entsprechend hoch.

Auch Support-Vektor-Modelle, als eine weitere Art von Kernel-Modellen, sind universelle Approximatoren. Der Modellausgang \hat{y} ergibt sich aus

$$\hat{y} = \Psi \hat{\Theta}. \quad (5.25)$$

Die Modellparameter werden dabei durch die Minimierung der Zielfunktion bestimmt

$$\hat{\Theta} = \arg \min_{\hat{\Theta}, \xi_i, \xi_i^*} \left(\frac{1}{2} \hat{\Theta}^T \hat{\Theta} + C \sum_{i=1}^N (\xi_i + \xi_i^*) \right) \quad (5.26)$$

mit den einzuhaltenden Randbedingungen

$$\begin{aligned} y_i - \Psi_i \hat{\Theta} &\leq \epsilon + \xi_i \\ \Psi_i \hat{\Theta} - y_i &\leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0 \end{aligned} \quad (5.27)$$

für alle Messdaten $i = 1, \dots, N$. Die Optimierung minimiert den quadratischen Fehler zwischen Messdaten und Modell bei gleichzeitiger Minimierung der Abstände der Ausreißer ξ_i und ξ_i^* unter Berücksichtigung, dass möglichst viele Messdaten innerhalb des 2ϵ -Toleranzbandes um \hat{y} liegen. Die Gewichtung C und die Toleranzbreite ϵ sind dabei manuell einzustellende Parameter. Nachteilig sind der hohe Rechenzeitbedarf, da bei der Modellbildung eine Optimierung sowie die notwendige heuristische Wahl der beiden Einstellparameter C und ϵ durchgeführt wird.

5.1.3 Modellbildungsverfahren für dynamische Prozesse

Lineare zeitinvariante dynamische Prozesse können mit verschiedenen Modellstrukturen und Parameterschätzmethoden identifiziert werden, siehe z.B. [36, 62, 63, 100, 150]. Die vorliegende Arbeit beschränkt sich, basierend auf der Klassifizierung nach [173], auf dynamische Modelle, welche linear parametrische Ein-/Ausgangsmodelle im diskreten Zeitbereich sind. Diese sind zur Identifikation gut geeignet, da sie trotz einer geringen Anzahl an Parametern das dynamische Prozessverhalten exakt beschreiben können [173]. Die Schätzung erfolgt dabei, basierend auf der Prädiktionsfehler-Methode aus Abschnitt 5.1.2 mit Ein-/Ausgangsmodellen in Form von z-Übertragungsfunktionen, wobei zwischen einer Minimierung des Ein-Schritt-Prädiktorfehlers in Gleichungsfehleranordnung (ARX-Modell) und Minimierung des simulierten Prädiktionsfehlers in Ausgangsfehleranordnung (OE-Modell) unterschieden wird. Auf die Beschreibung von weiteren Verfahren, welche im Rahmen der Parameterschätzung eingesetzt werden können (ARMAX, BJ, IV, FIR, etc.), sei an dieser Stelle auf die entsprechende Literatur verwiesen, wie z.B. [70, 173, 100].

Die betrachteten Modelle beschreiben einen linearen dynamischen Prozess ohne Totzeit und ohne Durchgriff durch eine lineare Differentialgleichung im kontinuierlichen Zeitbereich

$$c_0 y(t) + c_1 \dot{y}(t) + c_2 \ddot{y}(t) + \dots + c_C y^{(C)}(t) = d_0 + d_1 \dot{u}(t) + d_2 \ddot{u}(t) + \dots + d_D u^{(D)}(t) \quad (5.28)$$

bzw. nach Laplace-Transformation durch die s-Übertragungsfunktion

$$G(s) = \frac{y(s)}{u(s)} = \frac{d_0 + d_1 s + d_2 s^2 + \dots + d_D s^D}{c_0 + c_1 s + c_2 s^2 + \dots + c_C s^C}. \quad (5.29)$$

Durch Abtastung und z-Transformation mit $HG(z) = \frac{z-1}{z} \xi \left\{ \frac{G(s)}{s} \right\}$ folgen im diskreten Zeitbereich die z-Übertragungsfunktion zu

$$G(z) = \frac{y(z)}{u(z)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_D z^{-D}}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_C z^{-C}} \quad (5.30)$$

sowie die dazugehörige Differenzialgleichung zu

$$y(kT_0) + a_1 y((k-1)T_0) + a_2 y((k-2)T_0) + \dots + a_C y((k-C)T_0) = b_0 u(kT_0) + b_1 u((k-1)T_0) + b_2 u((k-2)T_0) + \dots + b_D u((k-D)T_0). \quad (5.31)$$

Die Stell- sowie gemessene Ausgangsgröße sind dabei um ihren Stationärwert gebildet:

$$\begin{aligned} u(kT_0) &= U(kT_0) - U_{00} \\ y(kT_0) &= Y(kT_0) - Y_{00} \end{aligned} \quad (5.32)$$

Bei der Schätzung der Parameter sowie Bestimmung der Modellausgangsgröße des dynamischen Modells muss zwischen der Gleichungsfehler- und Ausgangsfehleranordnung unterschieden werden. Die Regressormatrix für die Gleichungsfehleranordnung folgt zu

$$\Psi_{\text{ARX}} = \begin{bmatrix} u(0) & \cdots & u(C-1) & -y(0) & \cdots & -y(D-1) \\ u(1) & \cdots & u(C) & -y(1) & \cdots & -y(D) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u(N) & \cdots & u(C+N-1) & -y(N) & \cdots & -y(D+N-1) \end{bmatrix}. \quad (5.33)$$

Die Schätzung mit der Gleichungsfehleranordnung entspricht dabei einer Schätzung mit ARX-Modell, bei welchem die Störübertragungsfunktion nur eine Nennerdynamik aufweist [70]. Die Modellausgangsgröße wird mit einer Ein-Schritt-Prädiktion auf Basis der gemessenen Ausgangsgröße [173] bestimmt, die maximal ein Zeitschritt zurückliegt. Im Gegensatz dazu stehen in der Regressormatrix der Ausgangsfehleranordnung nicht die gemessenen Ausgangsgrößen, sondern die modellierte Ausgangsgröße:

$$\Psi_{\text{OE}} = \begin{bmatrix} u(0) & \cdots & u(C-1) & -\hat{y}(0) & \cdots & -\hat{y}(D-1) \\ u(1) & \cdots & u(C) & -\hat{y}(1) & \cdots & -\hat{y}(D) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u(N) & \cdots & u(C+N-1) & -\hat{y}(N) & \cdots & -\hat{y}(D+N-1) \end{bmatrix}. \quad (5.34)$$

Diese Schätzung entspricht der Schätzung mit OE-Modell, bei dem die Störübertragungsfunktion sowohl eine Zähler- als auch eine Nennerdynamik aufweist. Sie hat zur Folge, dass die Bestimmung der Parameter mit nichtlinearen Verfahren (aufgrund der Nichtlinearität des Modells in den Parametern) sowie die Bestimmung der Modellausgangsgröße iterativ für jeden Zeitschritt erfolgen muss. Vorteile der OE-Schätzung sind neben der besseren Validierungsgüte des Modells, welche sich aufgrund der konsistenten Schätzung der von der Störübertragungsfunktion unabhängigen Eingangsübertragungsfunktion ergibt, auch die geringe Abhängigkeit von der Abtastzeit T_0 . Dadurch ist die OE-Schätzung für Prozesse, die eine kleine Abtastzeit im Vergleich zur Einschwingzeit aufweisen, und somit auch für Prozesse mit stark unterschiedlichen Zeitkonstanten geeignet [173].

Zur Bestimmung des Modells müssen die dynamischen Ordnungen C und D geeignet gewählt werden. Bei der Betrachtung verschiedener dynamischer Übertragungsfunktionen (siehe Tabelle 5.1 bzw. [64]) werden verschiedenen Zusammenhänge deutlich, die bei der Wahl der dynamischen Ordnung zu beachten sind:

- Die Hinzunahme einer weiteren Zeitkonstante führt zu einer Erhöhung der Zähler- und Nennerordnung um 1.
- Der Einbezug eines D-Anteils hat dagegen keine Auswirkung auf die Ordnung, sondern ausschließlich auf die diskreten Parameter.

- Einen konstanten Zählerterm b_0 gibt es lediglich bei sprungförmigen Prozessen, bei denen die Zähler- und Nennerordnung gleich sind.
- Im Fall von MISO-Prozessen entspricht die diskrete Zähler- und Nennerordnung der Anzahl an Zeitkonstanten.

Zusammenfassend kann festgehalten werden, dass bei der Parameterschätzung dynamischer diskreter Modelle die dynamischen Ordnungen C und D gleich sowie mindestens so hoch wie die Anzahl an Prozesszeitkonstanten festgelegt werden müssen. Das Erhöhen der dynamischen Ordnung entspricht dem Hinzufügen einer weiteren Zeitkonstante in die Prozessgleichung.

Tabelle 5.1: Vergleich von s- und z-Übertragungsfunktionen verschiedener Prozesse zur Ermittlung der dynamischen Ordnung (Zeitverschiebung). Die Parameter sind dabei zeilenweise unterschiedlich.

Prozess	$G(s)u(s)$	$G(z)u(z)$
$PT_1 \cdot u_1$	$\frac{d_0}{c_0+c_1s}u_1$	$\frac{b_1z^{-1}}{a_0-a_1z^{-1}}u_1$
$PT_2 \cdot u_1$	$\frac{d_0}{c_0+c_1s+c_2s^2}u_1$	$\frac{b_1z^{-1}+b_2z^{-2}}{a_0-a_1z^{-1}+a_2z^{-2}}u_1$
$PT_3 \cdot u_1$	$\frac{d_0}{c_0+c_1s+c_2s^2+c_3s^3}u_1$	$\frac{b_1z^{-1}+b_2z^{-2}+b_3z^{-3}}{a_0-a_1z^{-1}+a_2z^{-2}-a_3z^{-3}}u_1$
$PD_2T_3 \cdot u_1$	$\frac{d_0+d_1s+d_2s^2}{c_0+c_1s+c_2s^2+c_3s^3}u_1$	$\frac{b_1z^{-1}+b_2z^{-2}+b_3z^{-3}}{a_0-a_1z^{-1}+a_2z^{-2}-a_3z^{-3}}u_1$
$PD_3T_3 \cdot u_1$	$\frac{d_0+d_1s+d_2s^2+d_3s^3}{c_0+c_1s+c_2s^2+c_3s^3}u_1$	$\frac{b_0+b_1z^{-1}+b_2z^{-2}+b_3z^{-3}}{a_0-a_1z^{-1}+a_2z^{-2}-a_3z^{-3}}u_1$
$\begin{pmatrix} PT_1 \\ PT_1 \end{pmatrix} (u_1, u_2)$	$\frac{(d_0+d_1s)u_1 + (d_2+d_3s)u_2}{c_0+c_1s+c_2s^2}$	$\frac{(b_1z^{-1}+b_2z^{-2})u_1 + (b_3z^{-1}+b_4z^{-2})u_2}{a_0-a_1z^{-1}+a_2z^{-2}}$
$\begin{pmatrix} PT_1 \\ PT_1 \\ PD_1T_1 \end{pmatrix} (u_1, u_2, u_3)$	$\frac{(d_0+d_1s+d_2s^2)u_1 + (d_3+d_4s+d_5s^2)u_2 + (d_6+d_7s+d_8s^2+d_9s^3)u_3}{c_0+c_1s+c_2s^2+c_3s^3}$	$\frac{(b_1z^{-1}+b_2z^{-2}+b_3z^{-3})u_1 + (b_4z^{-1}+b_5z^{-2}+b_6z^{-3})u_2 + (b_7+b_8z^{-1}+b_9z^{-2}+b_{10}z^{-3})u_3}{a_0-a_1z^{-1}+a_2z^{-2}+a_3z^{-3}}$
$\begin{pmatrix} PT_1 \\ PT_2 \\ PD_1T_2 \end{pmatrix} (u_1, u_2, u_3)$	$\frac{(d_0+d_1s+d_2s^2+d_3s^3+d_4s^4)u_1 + (d_5+d_6s+d_7s^2+d_8s^3)u_2 + (d_9+d_{10}s+d_{11}s^2+d_{12}s^3+d_{13}s^4)u_3}{c_0+c_1s+c_2s^2+c_3s^3+c_4s^4+c_5s^5}$	$\frac{(b_1z^{-1}+b_2z^{-2}+b_3z^{-3}+b_4z^{-4}+b_5z^{-5})u_1 + (b_6z^{-1}+b_7z^{-2}+b_8z^{-3}+b_9z^{-4}+b_{10}z^{-5})u_2 + (b_{11}z^{-1}+b_{12}z^{-2}+b_{13}z^{-3}+b_{14}z^{-4}+b_{15}z^{-5})u_3}{a_0-a_1z^{-1}+a_2z^{-2}+a_3z^{-3}+a_4z^{-4}+a_5z^{-5}}$

Ausgehend von Gleichung (5.31) kann das statische Modell bestimmt werden, indem die Gleichung für den stationären Fall betrachtet wird [90]. Dabei nehmen die zeitlich verzögerten Ein- und Ausgangsgrößen die gleichen Werte an:

$$\begin{aligned} y &= y(kT_0) = y((k-1)T_0) = y((k-2)T_0) = \dots = y((k-C)T_0) \\ u &= u(kT_0) = u((k-1)T_0) = u((k-2)T_0) = \dots = u((k-D)T_0) \end{aligned} \quad (5.35)$$

Zusammen mit Gleichung (5.31) folgt

$$y = \frac{(b_0 + b_1 + \dots + b_D)u}{(a_0 + a_1 + \dots + a_C)} \quad (5.36)$$

Der statische Parameter, welcher den linearen Zusammenhang zwischen y und u darstellt, wird dabei nach Gleichung 5.36 direkt aus den dynamischen Parametern bestimmt, ohne eine zusätzliche Schätzung der statischen Parameter durchführen zu müssen.

Eine Erweiterung der Parameterschätzung auf Modelle mit mehreren bzw. nichtlinear transformierten Eingangsgrößen erfolgt durch Erweiterung der Regressormatrix aus Gleichung (5.33) bzw. (5.34) um die entsprechenden Regressoren. Das statische Modell wird dann bei der NARX- bzw. NOE-Schätzung durch Bilden der Parametersumme der statisch zusammengehörenden Regressoren und Dividieren durch $(a_0 + a_1 + \dots + a_C)$ abgeleitet.

5.1.4 Verfahren zur Bewertung der Modellgüte

Im Folgenden werden verschiedene Bewertungskriterien vorgestellt. Dabei wird nur auf solche eingegangen, welche im Rahmen einer automatischen Vermessung anwendbar und direkt interpretierbar sind, d.h. entweder Kriterien mit der gleichen Einheit wie die zu modellierende Größe oder wie die prozentualen Kriterien. Grafische Kriterien, wie z.B. die in [91] Beschriebenen, sind ausschließlich für eine manuelle Bewertung und Analyse von Modellen geeignet und nicht für die Anwendung in einer automatisierten Vermessung.

Ein Modell wird mit verschiedenen Kriterien für den Trainings- oder den Validierungsfehler bewertet. Der Trainingsfehler ist der Fehler zwischen Prozess- und Modellausgang, welcher sich nach der Schätzung der Parameter für die zur Identifikation benutzten $(N - O)$ Messdaten ergibt:

$$J_{\text{SAE},t} = \sum_{i=1}^{N-O} |y_i(\mathbf{u}_i) - \hat{y}_i(\mathbf{u}_i)| \quad (5.37)$$

Wird der Prozessausgang für neue, nicht zur Parameterschätzung verwendete, O Messdaten simuliert, ergibt sich entsprechend der Validierungsfehler $J_{\text{SAE},v}$:

$$J_{\text{SAE},v} = \sum_{i=N-O+1}^N |y_i(\mathbf{u}_i) - \hat{y}_i(\mathbf{u}_i)| \quad (5.38)$$

Die Gütekennzahl J_{SAE} (summed absolute error) stellt dabei die Summe der Absolutwerte der Fehler und somit die Abweichung zwischen den Mess- und Schätzwerten dar, welche für ein Modell möglichst klein sein sollte. Sie eignet sich ausschließlich zum Vergleich von verschiedenen Modellen mit jeweils gleicher Anzahl an Trainings- bzw. Validierungsdaten. Weisen die zu vergleichenden Modelle eine unterschiedliche Anzahl an zugrundeliegenden Daten auf, sollte stattdessen die J_{MAE} Güte (mean absolute error) herangezogen werden

$$J_{\text{MAE}} = \frac{1}{N - O} \sum_{i=1}^{N-O} |y_i(\mathbf{u}_i) - \hat{y}_i(\mathbf{u}_i)|, \quad (5.39)$$

welche einen durchschnittlichen Modellfehler angibt. Eine weitere eingesetzte Kennzahl ist die Wurzel aus dem gemittelten quadratischen Fehler J_{RMSE} (root mean squared error):

$$J_{\text{RMSE}} = \sqrt{\frac{1}{N - O} \sum_{i=1}^{N-O} (y_i(\mathbf{u}_i) - \hat{y}_i(\mathbf{u}_i))^2} \quad (5.40)$$

Dabei ist zu berücksichtigen, dass durch das Quadrieren große Abweichungen stärker gewichtet werden als kleine, wodurch Ausreißer einen hohen Einfluss auf die Kennzahl haben [89]. Sollen Modelle mit verschiedenen Ausgangsgrößen verglichen werden, muss die Kennzahl normiert werden. Dabei können verschiedene Normierungen zum Einsatz kommen:

$$\begin{aligned}
 J_{\text{NRMSEr}} &= \frac{\sqrt{\frac{1}{N-O} \sum_{i=1}^{N-O} (y_i(\mathbf{u}_i) - \hat{y}_i(\mathbf{u}_i))^2}}{y_{\max} - y_{\min}} \\
 J_{\text{NRMSEm}} &= \frac{\sqrt{\frac{1}{N-O} \sum_{i=1}^{N-O} (y_i(\mathbf{u}_i) - \hat{y}_i(\mathbf{u}_i))^2}}{\bar{y}} \\
 J_{\text{NMSEt}} &= \frac{\frac{1}{N-O} \sum_{i=1}^{N-O} (y_i(\mathbf{u}_i) - \hat{y}_i(\mathbf{u}_i))^2}{\frac{1}{N-O} \sum_{i=1}^{N-O} (y_i(\mathbf{u}_i) - \bar{y})^2} = \frac{\frac{1}{N-O} \text{SSE}}{\frac{1}{N-O} \text{SST}}
 \end{aligned} \tag{5.41}$$

Neben der Normierung der RMSE-Kennzahl auf den Bereich (NRMSEr) bzw. den Mittelwert (NRMSEm) der Ausgangsgröße kann auch die in [89] vorgestellte Normierung der MSE-Kennzahl auf die mittlere Quadratsumme der Differenz aus Mess- und Mittelwert (NRMSEt) durchgeführt werden. Dadurch kann die Kennzahl als das Verhältnis der Summe der nicht erklärten Abweichungen (SSE) zur Summe der gesamten Abweichungen (SST) interpretiert werden. Eine gute Messdatenapproximation zeichnet sich dabei durch einen Wert nahe Null aus. Für das Verhältnis der Summe der erklärten Abweichungen zur Summe der gesamten Abweichungen bei der linearen Regression folgt

$$1 - J_{\text{NMSEt}} = R^2 = \frac{\sum_{i=1}^{N-O} (\hat{y}_i(\mathbf{u}_i) - \bar{y})^2}{\sum_{i=1}^{N-O} (y_i(\mathbf{u}_i) - \bar{y})^2} \tag{5.42}$$

Das Bestimmtheitsmaß R^2 ist bei der Regression ein Indikator für den linearen Zusammenhang zwischen dem Modell und den Messdaten [6]. Ein Wert nahe eins deutet dabei auf einen stark linearen Zusammenhang hin, ein Wert nahe null entsprechend auf keinen linearen Zusammenhang. Kein linearer Zusammenhang kann dabei entweder gar kein Zusammenhang zwischen Messdaten und Modell bedeuten, oder aber auch dass es einen nichtlinearen Zusammenhang gibt. Das Bestimmtheitsmaß ist nicht geeignet, um eine Aussage zwischen kein oder nicht-lineare Zusammenhang zu treffen. Zu beachten ist weiterhin, dass mit zunehmender Anzahl an Modellparametern das Bestimmtheitsmaß größer wird, wodurch das Modell überangepasst wird.

Ein großer Trainingsfehler ergibt sich bei einer Unteranpassung des Modells. Eine Erhöhung der Strukturkomplexität führt zu einer besseren Anpassung, woraus sich ein immer kleiner werdender Trainingsfehler ergibt. Dabei wird das Modell aber nicht nur an den Prozess, sondern auch an die Messstörungen angepasst. Ein sehr kleiner Trainingsfehler, d.h. eine gute Modellierung von Prozess und Störung, deutet auf eine zu komplexe Struktur und somit eine Überanpassung des Modells an die vorliegenden Messdaten hin. Dies wird in dem Validierungsfehler deutlich. Der Validierungsfehler sinkt, aufgrund des stark fallenden Bias-Fehlers, mit zunehmender Modellkomplexität bis er ein Minimum erreicht und steigt danach wieder an, aufgrund des immer größeren Varianz-Fehlers. Mit dem Trainingsfehler kann somit die generelle Eignung des

Modells zur Identifikation des Prozesses bewertet werden, während der Validierungsfehler die Wahl der Komplexität eines Modells bewertet.

Zur Bestimmung des Validierungsfehlers werden zusätzliche Messdaten benötigt, welche nicht zur Parameterschätzung verwendet wurden. Resamplingverfahren nutzen den bereits vorhandenen Datensatz mit $N - O$ Messdaten um daraus sowohl den Trainings- als auch den Validierungsfehler zu bestimmen, ohne die zusätzlichen O Messdaten erfassen zu müssen. Das Verfahren der Kreuzvalidierung [26] teilt den Datensatz in L möglichst gleichmäßig im Versuchsraum verteilte Messdatensätze. Ausgehend davon wird L mal eine Parameterschätzung für ein Modell fester Struktur durchgeführt, jeweils mit $(L - 1)$ Messdatensätzen als Trainingsdaten und einem Teildatensatz als Validierungsdaten. Die Validierungsgüte der Modellstruktur ergibt sich durch den Mittelwert aller Validierungsgüten der L Modelle:

$$J_{\text{SAE},v} = \sum_{j=1}^L J_{\text{SAE},v,j} \quad (5.43)$$

Das Jackknife-Verfahren [116] (auch Leave-One-Out-Verfahren genannt) ist der Sonderfall der Kreuzvalidierung, bei dem gilt $L = N - O$. Eine weitere Möglichkeit besteht darin, nacheinander $1, \dots, L$ einzelne Datenpunkte aus dem vollständigen Datensatz mit $(N - O)$ -Datenpunkten zu ziehen, wobei die einzelnen Datenpunkte immer wieder in den gesamten Datensatz zurückgelegt werden. Folglich liegen in der Stichprobe Datenpunkte doppelt vor, welche dann zur Modellbildung und der vollständige Datensatz zur Modellvalidierung genutzt werden können, da die Zusammensetzung beider unterschiedlich ist [89]. Bei diesem Bootstrap-Verfahren ergibt sich der Validierungsfehler durch das wiederholte Identifizieren mit verschiedenen Stichproben. Eine Übersicht der verschiedenen Verfahren ist in [68] zu finden.

5.1.5 Bias-Varianz Dilemma

Das Bias-Varianz Dilemma beschreibt den Zusammenhang zwischen der Komplexität und den Generalisierungseigenschaften eines Modells. Unter Annahme eines festen Datensatzes mit N Messdaten sowie weißem Rauschen als Messstörungen, und somit unkorrelierten Fehlern mit einem Erwartungswert $E\{e\} = 0$ und konstanter Varianz σ^2 , kann der Erwartungswert der Residuen in einen Erwartungswert des Modellfehlers und einen Erwartungswert der Varianz der Messstörung zerlegt werden:

$$E\{e^2\} = E\{(y - \hat{y})^2\} + E\{\sigma^2\}. \quad (5.44)$$

Der Erwartungswert des Modellfehlers kann wiederum in zwei weitere Fehlerterme zerlegt werden [70]:

$$E\{(y - \hat{y})^2\} = E\{[y - E\{\hat{y}\}]^2\} + E\{[\hat{y} - E\{\hat{y}\}]^2\} = \text{Bias}^2 + \text{Varianz}. \quad (5.45)$$

Der systematische Bias-Fehler ist dabei der Fehler, welcher sich aufgrund der Modellunflexibilität, basierend auf einer zu geringen Komplexität der Modellstruktur, ableitet. Die dabei entstehende Unteranpassung des Modells hat zwar einen großen systematischen Modellfehler, ist

jedoch unempfindlich gegenüber Änderungen der Messdaten. Die Empfindlichkeit gegenüber den Trainingsdaten wird über den zweiten Term, den Varianz-Fehler, bewertet. Eine zu hohe Komplexität des Modells führt zu einer starken Flexibilität und somit zur Gefahr der Überanpassung des Modells an die Messdaten mitsamt des Messrauschens.

Dieser Zielkonflikt, eine möglichst geeignete Modellstruktur zu wählen, welche weder eine Über- noch eine Unteranpassung zur Folge hat, wird als das Bias-Varianz-Dilemma bezeichnet [120, 70]. Zur Lösung der Bias-Varianz-Dilemmas muss mithilfe von Verfahren zur Strukturbestimmung sowie Resamplingverfahren zur Validierung von Modellen eine optimale, d.h. für die N Messdaten geeignete, Modellstruktur bestimmt werden, welche den tatsächlichen Modellfehler minimiert.

5.1.6 Strukturbestimmung von Modellen

Eine geeignete Modellstruktur wird über einen geeigneten Auswahlalgorithmus bestimmt. Dabei wird aus einem Kandidatensatz mit allen Q potentiellen Regressoren ein Teilmenge mit R Regressoren bestimmt, welche den Prozess am besten beschreiben. Die Teilmenge wird dabei anhand eines geeigneten Informationskriteriums bewertet.

Informationskriterien

Zur Bewertung können verschiedene Kriterien zum Einsatz kommen. *Mallows' C_p -Statistik* [23] kombiniert den auf die Varianz standardisierten quadratischen Fehler (SE) mit einem Strafterm für die Anzahl N an Messpunkten und Anzahl R an ausgewählten Modellregressoren zu

$$C_p = \frac{\sum_{i=1}^N (y(i) - \hat{y}_R(i))^2}{\hat{\sigma}^2} - N + 2p \quad \text{mit } p = (R + 1). \quad (5.46)$$

Dabei sind y die gemessene Ausgangsgröße und \hat{y}_R der mit den R Regressoren abgeleitete Modellausgang. Die Varianz $\hat{\sigma}^2$ wird aus der Quadratfehlersumme des Modells, welches mit allen Regressoren Q bestimmt wurde, und der Anzahl an Freiheitsgraden (N_Q) berechnet:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N (y(i) - \hat{y}_Q(i))^2}{N - Q} \quad (5.47)$$

Unter der Voraussetzung von mittelwertfreien, unabhängigen und identisch verteilten Fehlern, kann mit C_p die Regressorauswahl bewertet werden. Der erste Term sinkt dabei mit einer zunehmenden Anzahl an Modellparametern (sinkender Bias-Fehler), während der zweite Term mit der Anzahl an Modellparametern und Messpunkten steigt (steigender Varianz-Fehler). Der Gesamtfehler fällt folglich mit einem kleiner werdenden C_p -Wert ab.

Im Rahmen der Online-Methodik ist speziell zu Beginn der Vermessung die Anzahl N an Messdaten noch gering im Vergleich zu der Anzahl Q an potentiellen Regressoren. Um bereits zu diesem Zeitpunkt eine Auswahl zu ermöglichen, kann die angepasste \bar{C}_p -Statistik [43] angewandt werden, für die gilt:

$$\bar{C}_p = C_p - \frac{2(Q - R + 1)}{N - Q - 3} \quad (5.48)$$

Ein weiteres Informationskriterium, welches zur Modellauswahl auch bei wenigen Messpunkten verwendet werden kann, ist das korrigierte *Akaike's information criterion* (AICc, [23]). Dabei wird nicht die Güte des Modells als Absolutwert bewertet, sondern die Wahrscheinlichkeit, den Informationsverlust durch das jeweilige Modell zu minimieren. Demzufolge hat das Modell mit einem möglichst kleinem Kriterium die größte Wahrscheinlichkeit, die Messdaten darzustellen. Für das Kriterium gilt:

$$\text{AICc} = 2R - 2 \ln(\mathcal{L}) + \frac{2R(R+1)}{N-R-1} \quad (5.49)$$

Hierbei ist $\ln(\mathcal{L})$ die logarithmierte Likelihood-Funktion des Modells, welche im Fall der Least-Squares-Schätzung (unabhängige, normalverteilte Residuen mit $\mu = 0$) zu

$$\ln(\mathcal{L}) = -\frac{N}{2} [\ln(2\pi\hat{\sigma}^2) + 1] = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln \left(\frac{\sum_{i=1}^N (y(i) - \hat{y}(i))^2}{N} \right) - \frac{N}{2} \quad (5.50)$$

folgt. Weiterhin steht für R die Anzahl an effektiven Parametern des Modells sowie N für die Anzahl an Messpunkten. Im Vergleich zur C_p -Statistik ist das korrigierte Kriterium AICc zur Modellbewertung bei wenig Messpunkten im Vergleich zur Anzahl an Modellparametern geeignet und generell einsetzbar solange die Wahrscheinlichkeitsverteilung einer Normalverteilung entspricht. Für $N \rightarrow \infty$ nähert sich das korrigierte dem normalen AIC an [106].

Weitere Informationskriterien die in der Modellauswahl zum Einsatz kommen sind z.B. das Schwarz BIC [152], welches die Anzahl der Parameter nicht mit einem konstanten Faktor, sondern in Abhängigkeit von der Anzahl an Messpunkten bestraft, sowie das korrigierte Bestimmtheitsmaß \overline{R}^2 [158], welches das normale Bestimmtheitsmaß um eine Bestrafung der Modellkomplexität erweitert.

Alle vorgestellten Informationskriterien basieren auf der Trainingsgüte. Eine weitere Möglichkeit besteht im Einsatz einer Kreuzvalidierung. Dabei werden die N Messdaten in L Teilmengen unterteilt. Daraufhin wird mit jeweils $L - 1$ Teilmengen das Modelltraining und mit dem verbleibenden Messdatensatz die Validierung durchgeführt. Dieses Vorgehen wird für jeden der L Datensätze wiederholt und die mittlere Validierungsgüte oder das Pseudo-Bestimmtheitsmaß PRESS [116] bestimmt. Wesentlicher Vorteil der Kreuzvalidierung ist, dass sie auf einer Validierungsgüte basiert, wodurch das tatsächliche Minimum des Bias-Varianz-Dilemmas gefunden werden kann. Nachteil ist der hohe Rechenaufwand, wodurch diese für einen Einsatz im Rahmen der Online-Methode ungeeignet ist.

Neben der Bewertung mit Informationskriterien kann auch eine Bewertung mithilfe von Hypothesentests erfolgen [123, 14]. Dabei wird, basierend auf einer studentischen t -Verteilung, die Hypothese, dass ein Parameter aus dem Modell entfernt werden soll, getestet. Hierdurch kann durch Vorgabe einer Signifikanzgrenze für die Wahrscheinlichkeit $t_{1-\alpha}$ die Parameterauswahl bewertet werden.

Auswahlalgorithmen

Die Bestimmung der für ein Modell relevanten Regressoren kann durch die Parameterschätzung und Gütebestimmung der Modelle für alle möglichen Kombinationen der Q Regressoren

erfolgen. Dieses Vorgehen ist bereits für die hohe Anzahl an möglichen Modellen bei kleinen Modellordnungen S sowie geringer Anzahl an Stellgrößen P nicht praktikabel. Aus diesem Grund werden Austauschalgorithmen eingesetzt.

Zu den einfachsten Algorithmen gehörten die Methoden der *schrittweisen Regression*. Neben reinen Vorwärts-, Rückwärts- oder Austauschalgorithmen [106] kann auch eine kombinierte Strategie aller drei Algorithmen eingesetzt werden [146]. Dabei wird im Anschluss an eine Vorwärtswahl die Rückwärtswahl der Regressoren solange durchgeführt bis die Modellgüte nicht mehr ansteigt. Anschließend wird ein Regressoraustausch durchgeführt, um Regressoren, welche einen höheren Einfluss als die bisher in der Teilmenge vorhandenen, einzubeziehen. Vorteil dieses Algorithmus ist die einfache Umsetzbarkeit unabhängig von dem Modelltyp. Nachteil ist, dass der heuristische Algorithmus zwar ein besseres Optimum als die einzelnen Vorwärts-, Rückwärts- oder Austauschalgorithmen findet, jedoch kein globales Optimum. Weiterhin ist die Bestimmung zeitaufwendig und stark von der Anzahl an Modellordnungen S und Stellgrößen P abhängig, da nach jedem einzelnen Schritt eine neue Identifikation zur Bewertung der Modellgüte durchgeführt werden muss.

Neben den heuristischen schrittweisen Auswahlverfahren gibt es auch datenbasierte Verfahren, welche mithilfe von statistischen Methoden angewandt auf die vorliegenden Daten die relevanten Regressoren bestimmen. Die *Hauptkomponentenanalyse* (engl. principal component analysis [76, 72]) nutzt die Kovarianz-Matrix der Messdaten um aus einer Singulärwertzerlegung orthogonale Regressoren aus Linearkombinationen der ursprünglichen Regressoren zu bilden. Hierdurch wird die Anzahl an Regressoren reduziert, wodurch dieses Verfahren schneller als eine schrittweise Regression ist. Nachteil ist, dass durch die Kombination der Regressoren die ursprüngliche Struktur des Modells verloren geht und alle, nach der Linearkombination entstehenden, Koeffizienten in die Struktur einbezogen werden.

Die bisherigen Verfahren bestimmen die relevanten Regressoren heuristisch oder basierend auf der Kovarianz der Daten. Eine weitere Möglichkeit ist die Integration der Auswahl der relevanten Regressoren in die Parameterschätzung. Die *ridge regression* [61], welche einer Regression nach Tikhonov [160] entspricht, wird die LS-Schätzung eines Modells, welches linear in seinen Parametern ist, um einen Strafterm für die Parameter (euklidische Norm) zu erweitern:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg \min_{\boldsymbol{\theta}} \left(\sum_{i=1}^N \left(y_i(\mathbf{u}_i) - \boldsymbol{\Psi}_i(\mathbf{u}_i) \hat{\boldsymbol{\theta}} - b_0 \right)^2 + \lambda \sum_{j=1}^Q \hat{\theta}_j^2 \right) \\ &= (\boldsymbol{\Psi}^T \boldsymbol{\Psi} + \lambda \mathbf{I})^{-1} \boldsymbol{\Psi}^T \mathbf{y}.\end{aligned}\quad (5.51)$$

Zu berücksichtigen ist hierbei, dass der Strafterm ausschließlich die variablen Regressoren und nicht den konstanten Offset b_0 beinhaltet. Zudem sind alle Regressoren standardisiert (Varianz=1 und Mittelwert=0) sowie die Ausgangsgröße zentriert, um eine gleichmäßige Bestrafung der Koeffizienten zu erreichen. Die Regressoren müssen hierfür vor der Schätzung standardi-

siert werden:

$$\begin{aligned}\frac{1}{N} \sum x_{i,j} &= 0 \\ \frac{1}{N} \sum x_{i,j}^2 &= 1\end{aligned}\tag{5.52}$$

Die so erweiterte Schätzung entspricht für $\lambda = 0$ der klassischen Schätzung nach der Methode der kleinsten Quadrate. Mit dem Tuningparameter λ kann für Werte $\inf > \lambda > 0$ die Stärke der Parameterunterdrückung bestimmt werden. Dabei wird die Varianz reduziert, indem ein Bias eingeführt wird. Dem Vorteil der schnellen und in die Schätzung integrierten Bestimmung der unterdrückten Parameter steht der Nachteil gegenüber, dass lediglich eine Unterdrückung erfolgt, ohne dass die Parameter zu null gesetzt werden. Somit kann die Regressorauswahl nur durch Vorgabe eines Grenzwertes für die minimale Größe der Parameter erreicht werden. Weiterhin ist die Wahl des Tuningparameters heuristisch und stark vom Prozess abhängig.

Die Vorgabe eines Grenzwertes für die Parameter kann vermieden werden, indem der Strafterm die Parameter linear bestraft:

$$\hat{\Theta} = \arg \min_{\Theta} \left(\sum_{i=1}^N \left(y_i(u_i) - \Psi_i(u_i) \Theta - b_0 \right)^2 + \lambda \sum_{j=1}^Q \left| \hat{\Theta}_j \right| \right)\tag{5.53}$$

Hierbei sind $\lambda > 0$ der Gewichtungssparameter, $i = 1, \dots, N$ die Datenpunkte, $j = 0, \dots, Q$ die potentiellen Modellparameter sowie Ψ die Regressormatrix und \hat{y} der Modellausgang. Dieses Verfahren wird als *least absolute shrinkage and selection operator* (LASSO [159]) Verfahren bezeichnet. Der LASSO-Algorithmus erweitert die Kleinst-Quadrate-Schätzung der Modellparameter (erster Term) um die Summe der Beträge der Parameter (zweiter Term) als Bestrafungsterm, wodurch der Strafterm die Parameter linear bestraft. Im Vergleich zur *ridge regression* werden die Parameter um einen konstanten Betrag, anstelle einer proportionalen Verkleinerung, verringert, wodurch die Koeffizienten auch zu Null werden. Dies hat zur Folge, dass keine heuristische Grenze für den Wert der Parameter vorgegeben werden muss. Nachteilig ist jedoch, dass die Lösung von (5.53) nicht mehr analytisch möglich ist. Ein effizienter Algorithmus zur Bestimmung der LASSO Lösung ist der LARS-Algorithmus (*least angle regression* [31]), welcher genauso schnell ist wie eine schrittweise Vorwärts-Regression.

5.2 Online-Bias-Varianz-Dilemma

Identifikationsmethoden werden durch das Bias-Varianz-Dilemma charakterisiert. Dies besagt, dass ein Modell mit geringer Komplexität zur Unteranpassung führt (hoher Bias-Fehler), welche sich in einem hohen Trainings- und Validierungsfehler bemerkbar macht. Ein Modell mit hoher Komplexität führt dagegen zur Überanpassung (hoher Varianz-Fehler), was zu einem kleinen Trainingsfehler aber einem großen Validierungsfehler führt. Ziel bei der Modellbildung ist es, eine geeignete Struktur und somit Komplexität zu bestimmen, bei der der Gesamtfehler am kleinsten ist. Dies wird durch die Wahl einer geeigneten statischen bzw. dynamischen Ordnung, der Wechselwirkung sowie der Netzmodellteilung erreicht.

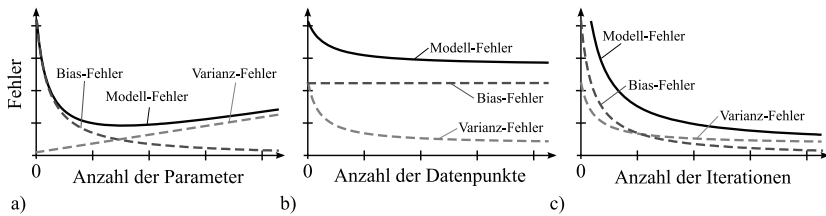


Bild 5.4: Qualitativer Vergleich des Bias-Varianz Dilemmas für die Annahmen a) konstanter Datensatz und variable Modellstruktur bzw. Parameteranzahl, b) variabler Datensatz und konstante Modellstruktur sowie c) variabler Datensatz und variable Modellstruktur bzw. Parameteranzahl. Simulative Beispiele für die Fälle b) und c) sind in den Abschnitten 4.3.6 und 5.4.6 zu finden.

Unter der Annahme einer konstanten Datensatzgröße kann der sich aus verschiedenen Modellkomplexitäten ergebende kleinste Gesamtfehler, bestehend aus Bias und Varianz, ermittelt werden (siehe Bild 5.4 a)), indem eine optimale Modellstruktur bestimmt wird.

Wird eine feste Struktur angenommen, verringert sich der Varianzfehler asymptotisch mit zunehmender Datensatzgröße auf einen konstanten Endwert, während der Bias-Fehler aufgrund der festen Modellstruktur konstant bleibt (siehe Bild 5.4 b)).

Im Rahmen der Online-Methodik erfolgt sowohl die Zunahme der Datensatzgröße als auch die Anpassung der Modellkomplexität iterativ. Dies hat zur Folge, dass sowohl der Bias- als auch der Varianzfehler reduziert werden, wodurch sich wiederum eine maximal mögliche, kontinuierliche Verringerung des Gesamtfehlers ergibt (siehe Bild 5.4 c)). Dadurch muss nicht, wie in klassischen Offline-Verfahren, der beste Zielkonflikt zwischen Bias und Varianz für einen Datensatz ermittelt werden, sondern die Struktur kann zielgerichtet unter Einbezug der zunehmenden Datensatzgröße angepasst werden.

Um die erreichbare Modellgüte zu erhöhen, muss die Modellstruktur aktiv, d.h. basierend auf den Messdaten, angepasst werden. Die Modellstruktur definiert sich bei Polynommodellen ausschließlich aus den Regressoren. Diese ergeben sich aus der statischen Ordnung (u^1, u^2, u^3, \dots) und der dynamischen Ordnung ($u(k-1), u(k-2), u(k-3), \dots$) der Stellgrößen sowie der Anzahl an Wechselwirkungen bei mehreren Stellgrößen ($u_1 u_2, u_1 u_2 u_3, u_{12} u_2, \dots$). Bei Netzmodellen wird die Modellstruktur zusätzlich über die Modellteilung bestimmt (z.B. Anzahl an lokal linearen Teilmodellen eines LOLIMOT [120]).

5.3 Online-Modellanalyse

Offline identifizierte Modellen werden mit normierten Gütekriterien nach Abschnitt 5.1.4 analysiert. Bei der Analyse von iterativ identifizierten Online-Modellen muss berücksichtigt werden, dass ein Modellvergleich, basierend auf den normierten J_{NRMSE} -Kriterien nach Gleichung (5.1.4) oder dem Bestimmtheitsmaß R^2 nach Gleichung (5.1.4), nicht zulässig ist. Der Grund hierfür

liegt in dem sich mit jedem neuen Messdatenpunkt ändernden Mittelwert der Ausgangsgröße \bar{y} und den sich damit auch ändernden Normierungsfaktoren. Weiterhin ist der Einsatz von absoluten Kriterien wie dem J_{SAE} (siehe Gleichung (5.43)) nicht möglich. Bei diesem wird die Summe des Modellfehlers über alle Messdatenpunkte gebildet, wodurch die Anzahl N an Messpunkten einen Einfluss auf das Kriterium hat. Mit jedem neuen Messdatenpunkt wird der Fehler, unter der Annahme von gleich bleibenden Fehlern für die alten Messdatenpunkte, größer, woraus folgt:

$$\lim_{N \rightarrow \infty} J_{SAE} = \infty \quad (5.54)$$

Sollen mit der Online-Methodik identifizierte Modelle aus verschiedenen Iterationen miteinander verglichen werden, können nur gemittelte, nicht normierte Kriterien wie J_{RMSE} oder J_{MSE} zum Einsatz kommen. Speziell für Netzmodelle wird dabei zwischen einem globalen, einem global-lokalen sowie einem lokalen Fehler unterschieden:

$$J_{RMSE,global} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (5.55)$$

$$J_{RMSE,lokal} = \sqrt{\frac{\sum_{i=1}^O (y_i - \hat{y}_{j,i})^2}{O}} \quad (5.56)$$

$$J_{RMSE,global-lokal} = \sqrt{\frac{\sum_{i=1}^O (y_i - \hat{y}_i)^2}{O}} \quad (5.57)$$

Hierbei sind N die Anzahl an insgesamt vorhandenen Messdatenpunkten und O entsprechend die Anzahl der Messdatenpunkte im betrachteten Stellbereich des lokalen Modells. Dabei bestimmt das lokale Kriterium den Fehler des lokalen Netzmodells \hat{y}_j innerhalb des lokalen Stellbereichs, während das global-lokale Kriterium den Fehler des globalen Netzmodells \hat{y} innerhalb des lokalen Stellbereichs festlegt. In beiden Fällen werden jeweils nur die in dem lokalen Stellbereich vorhandenen O Messdatenpunkte einbezogen. Das globale Kriterium bewertet dagegen das globale Modell im gesamten Stellbereich basierend auf allen vorhandenen Messdatenpunkten.

Zusätzlich muss bei einem Gütekriterium berücksichtigt werden, ob dieses Messdaten eines Trainings oder Messdaten einer Validierung bewertet. Bei der Bewertung des Trainings mit

$$J_{RMSE,global,t} = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad (5.58)$$

wird die Güte des Modells zur Bestimmung der N für das Training genutzten Messdatenpunkte bestimmt, während mit

$$J_{RMSE,global,v} = \sqrt{\frac{\sum_{i=1}^H (y_i - \hat{y}_i)^2}{H}} \quad (5.59)$$

die Güte bei Bestimmung von H unabhängigen (also nicht für das Training des Modells verwendeten) Validierungsmessdatenpunkten folgt.

5.4 Online lokal polynomiale Netzmodelle

Die im Abschnitt 5.1.2 beschriebenen Verfahren zur Offline-Modellbildung sind für die Identifikation von Modellen basierend auf festen Messdatensätzen entworfen worden. Im Rahmen einer Online-Vermessung werden, ausgehend von wenigen Messdaten, iterativ neue Messdaten erfasst und dem zur Identifikation genutzten Datensatz hinzugefügt. Die Forderung nach einer fortwährenden Anregung bei der dynamischen Prozessidentifikation [70] ermöglicht den Einsatz von Offline-Methoden im Rahmen einer Online-Vermessung solange die Modellbildung nicht länger dauert als die Messung.

Sowohl für die Modellierung von nichtlinearen statischen Modellen als auch für die Bestimmung von dynamischen Verbrennungsmotormodellen, müssen mehrere hundert bis zu mehrere tausend Messdatensamples erfasst und zur Identifikation genutzt werden. Der Einsatz von Offline-Methoden zur Modellbildung resultiert dadurch in große Modellierungsdauern, welche ab einem von der Vermessung abhängigen Zeitpunkt größer als die eigentliche Messdauer sind. Der Prüfstand müsste dann mit einer weiteren Messung warten, wodurch Prüfstandsbetriebszeit sinnlos verloren gehen würde. Zudem muss bei der dynamischen Vermessung darauf geachtet werden, dass keine langen Stillstandszeiten zwischen den einzelnen Iterationen auftreten, um konstante Zeilen in der Regressormatrix zu vermeiden.

Durch die Anpassung der Modellbildung für Online-Methoden kann der iterativ größer werdende Datensatz berücksichtigt werden. Dabei wird, durch den Einsatz von rekursiven Methoden für die Parameterschätzung und die Modellstrukturanpassung sowie durch Methoden zur Online-Modellanalyse, das Identifizieren eines vollständig neuen Modells in jeder Iteration vermieden. Das Verfahren zur Modellierung mit *online lokalen polynomialen Netzmodellen* (ONLOPOMOT) ermöglicht dadurch eine schnellere Identifikation von Modellen bei gleichzeitig zielgerichteter Online-Interaktion mit der Vermessung am Prüfstand.

5.4.1 Online-Adaption der Modellparameter

Basierend auf iterativ hinzukommenden Messdaten werden die Parameter eines Modells neu geschätzt. Im Fall der Schätzung mit dem LASSO-Verfahren wird hierfür der Gewichtungsparemeter λ zur Bestimmung des LASSO-Lösung (siehe Gleichung (5.53)) sowie die vollständige Netzmodellstruktur des alten Modells übernommen. Aufgrund des dem LASSO-Verfahren zugrundeliegenden LARS-Optimierungsverfahrens kann ein direktes Update der Parameter, wie es z.B. beim RLS-Verfahren der Fall ist, nicht durchgeführt werden. Mit der gleichbleibenden Gewichtung der Bestrafung der Parameter sowie gleichen Netzmodellstruktur wird jedoch sichergestellt, dass sich bei der Adaption der Modellparameter die Modellkomplexität nicht signifikant ändert.

5.4.2 Online-Adaption der Regressoren

Unter der Annahme einer festen Datensatzgröße kann das in Bild 5.4 a) dargestellte Bias-Varianz-Dilemma mithilfe von Informationskriterien, wie z.B. Mallows' C_p -Statistik oder dem Akaikes' korrigiertem Informationskriterium, gelöst werden.

In der Online-Vermessung werden jedoch iterativ neue Daten hinzugenommen, was dazu führt, dass das Ergebnis der Auswahl (und dadurch die Modellstruktur) der vorherigen Iteration nicht mehr zwangsläufig gültig ist. Die bei wenigen Messdaten ignorierten bzw. dominierenden Regressoren können bei größer werdendem Datensatz dominant bzw. irrelevant werden. Weiterhin ist die Auswahl basierend auf Mallows' C_p -Statistik nur für ein offline Verfahren mit konstanter Anzahl an Messpunkten geeignet. Die Anzahl an Messpunkten geht direkt in die Berechnung mit ein, wodurch eine Zunahme an Messpunkten automatisch zu einem kleiner werdenden C_p -Wert führt. Demzufolge müsste ein schrittweises Auswahlverfahren in jeder Iteration komplett neu initiiert werden.

Im Gegensatz zur Offline-Identifikation, wird bei der Online-Identifikation die Modellgüte sowohl durch die steigenden Anzahl an Messdaten als auch durch die Anpassung der Modellstruktur höchstmöglich gesteigert (siehe Bild 5.4), indem Bias- und Varianzfehler gleichzeitig reduziert werden. In der Online-Methodik muss somit nicht das Dilemma zwischen Bias- und Varianz gelöst werden, sondern vielmehr muss die Modellstruktur zielorientiert unter Berücksichtigung der steigenden Anzahl an Datenpunkten adaptiert werden.

Die Adaption erfolgt mit den hinzukommenden Messdaten integriert in die Parameterschätzung, da eine heuristische Auswahl (z.B. schrittweise) in jeder Iteration neu initiiert werden müsste und dadurch zeitaufwändig wäre. Die signifikantesten Regressoren werden dabei durch die Kombination des LASSO-Algorithmus (siehe Gleichung (5.53)) mit dem korrigierten AICc (siehe Gleichung (5.49)) ausgewählt. Die Schätzung mit dem LASSO-Algorithmus führt zu einer Regressorauswahl, da die lineare Bestrafung die Parameter von Regressoren ohne großen Einfluss zu 0 bestimmt. Der Gewichtungsparemeter λ wird bei der Online-Methodik automatisch gewählt, indem für den innerhalb einer Iteration konstanten Datensatz für verschiedene λ die sich ergebenden Modelle mithilfe des korrigierten AICc bewertet werden. Dadurch kann bereits bei wenigen Messpunkten mit der Auswahl gestartet werden.

Die Parameterbestimmung wird dabei in jeder Iteration neu initiiert und nicht rekursiv mit einem Vergessensfaktor durchgeführt. Hintergrund ist, dass durch die gezielte Vermessung in Modellbereichen mit geringer Prozessinformation sowie die Adaption der Modellstruktur die bisher bestimmten Parameter zur neuen Modellbildung nicht mehr geeignet sind.

5.4.3 Online-Adaption der Netzmodellstruktur

Werden Netzmodelle zur Identifikation eingesetzt, wird die Modellstruktur neben den Regressoren zusätzlich über die Modellteilung bestimmt. Bekannte Methoden [119, 146] nutzen dabei einen Algorithmus zur Modellpartition, welcher den gesamten Regressorraum in kleinere Hy-

perquader teilt, indem iterativ der lokale Hyperquader mit dem schlechtesten lokalen Modell mit einem festen Verhältnis geteilt wird.

Initiales Netzmodell

Diese Offline-Methoden beginnen die Identifikation mit dem gesamten Datensatz bestehend aus N Messdaten, wodurch der initiale Hyperquader über den gesamten relevanten Stellbereich bzw. den sich daraus ergebenden Bereich der Regressoren \mathbf{x}_{\min} und \mathbf{x}_{\max} mit

$$\begin{aligned}\mathbf{x}_{\min} &= \left[\min_{i=1,\dots,N} (x_{1,i}), \dots, \min_{i=1,\dots,N} (x_{R-1,i}) \right] \\ \mathbf{x}_{\max} &= \left[\max_{i=1,\dots,N} (x_{1,i}), \dots, \max_{i=1,\dots,N} (x_{R-1,i}) \right]\end{aligned}\quad (5.60)$$

aufgespannt wird. Das sich ergebende Hypervolumen des Prozesses, welches dem Produkt der Bereiche aller Regressoren entspricht, folgt zu:

$$V_{\text{init,Prozess}} = \prod_{o=1}^{R-1} (x_{o,\max} - x_{o,\min}) \quad (5.61)$$

Im Gegensatz dazu sind im Fall der Online-Vermessung in der ersten Iteration nur wenige aus Messdaten bestimmte Regressoren \mathbf{x}_l (mit $l = 1, \dots, O$) zur ersten Identifikation vorhanden. Für die O initialen Regressoren gilt dabei, dass je größer die Entfernung von \mathbf{x}_l zum Rand des Hyperquaders aus allen Messdaten und somit

$$\begin{aligned}\mathbf{x}_{\text{RS},\min} &= \left[\min_{l=1,\dots,O} (x_{1,l}), \dots, \min_{l=1,\dots,O} (x_{R-1,l}) \right] \gg \mathbf{x}_{\min} \\ \mathbf{x}_{\text{RS},\max} &= \left[\max_{l=1,\dots,O} (x_{1,l}), \dots, \max_{l=1,\dots,O} (x_{R-1,l}) \right] \ll \mathbf{x}_{\max}\end{aligned}\quad (5.62)$$

ist, desto kleiner ist das initial aufgespannte Hypervolumen $V_{\text{init,online}}$ des ersten lokalen Modells im Rahmen der Online-Vermessung:

$$\begin{aligned}V_{\text{init,Prozess}} \gg V_{\text{init,online}} &= \prod_{o=1}^{R-1} \Delta_{\text{RS},o} \\ &= \prod_{o=1}^{R-1} (x_{\text{RS},o,\max} - x_{\text{RS},o,\min})\end{aligned}\quad (5.63)$$

Hieraus folgt, dass der Regressorbereich des Prozesses wesentlich größer ist als der Regressorbereich des Online-Modells. In Bild 5.5 ist bei der 1. Iteration dieser Zusammenhang qualitativ für ein Modell mit zwei Regressoren ($R = 3$) dargestellt.

Das Aufspannen des initialen Hyperquaders auf das kleinere Hypervolumen $V_{\text{init,online}}$ führt zu einem globalen Modell mit kleinem Interpolations- und großem Extrapolationsbereich zum Ende der Online-Vermessung, wodurch die erreichbare Modellgüte signifikant reduziert wird. Zur Vermeidung dessen wird im Rahmen der Online-Vermessung der initiale Hyperquader nicht

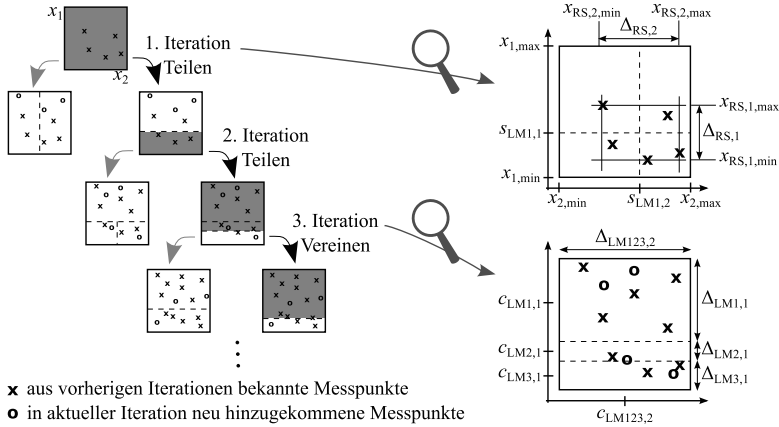


Bild 5.5: Achsen-orthogonaler Teilungs- und Vereinigungsalgorithmus zur Bestimmung der Netzmodellteilung eines ONLOPOMOT.

basierend auf den sich aus den initialen Messdaten ergebenden Regressoren aufgespannt. Vielmehr wird, basierend auf den minimalen und maximalen Regressoren

$$\begin{aligned} x_{RS,o,min} &\in \{u_{1,min} \dots, u_{P,min}, u_{1,min}^2, \dots, u_{P,min}^2, u_{1,min}u_{2,min}, \dots, u_{1,min}u_{P,min}\} \\ x_{RS,o,max} &\in \{u_{1,max} \dots, u_{P,max}, u_{1,max}^2, \dots, u_{P,max}^2, u_{1,max}u_{2,max}, \dots, u_{1,max}u_{P,max}\} \end{aligned} \quad (5.64)$$

welche aus den Limits der gemessenen physikalischen Größen u_{min} und u_{max} folgen, der initiale Hyperquader bereits zu Beginn der iterativen Vermessung als der maximal möglich vorkommender Hyperquader definiert. Hierdurch kann sichergestellt werden, dass der Extrapolationsbereich im Rahmen der Online-Vermessung klein bleibt.

Netzmodellteilung

Für den Einsatz im Rahmen der Online-Methodik muss die Modellstruktur zielgerichtet, unter Berücksichtigung der zunehmenden Anzahl an Messdaten, adaptiert werden. Die Bestimmung des Hyperquaders, welcher als nächstes geteilt wird, erfolgt mit dem im lokalen Stellbereich gemittelte Fehlerkriterium $J_{RMSE,lokal}$, welches zusätzlich um die Berücksichtigung des Volumens (J_{NRMSEV}) entsprechend Gleichung (4.30) bzw. der Anzahl an Messpunkten (J_{NRMSEs}) entsprechend Gleichung (4.31) erweitert wird. Bei der Adaption der Netzmodellteilung kann dadurch neben dem Modellfehler auch die Größe der lokalen Hyperquader bzw. die Anzahl an Messpunkten innerhalb der lokalen Hyperquader berücksichtigt werden.

Das Fehlerkriterium wird dabei an die aktuelle Art der Versuchsplanung gekoppelt. Hierfür wird die Annahme getroffen, dass das Hinzufügen einer weiteren Netzmodellteilung eine Reduktion des Trainingsfehlers zur Folge hat (sinkender Bias-Fehler). Erfolgt die Versuchsplanung raumfüllend, werden neue Messpunkte gleichmäßig über den gesamten Stellbereich verteilt. In

diesem Fall kommt das volumen-bezogene Kriterium zum Einsatz, welches die lokalen Hyper-volumen berücksichtigt. Die Fehlerreduktion durch die Teilung erfolgt dabei in Bereichen mit großen Fehlern und großen Volumen bzw. Modellstellbereichen, wodurch eine stärkere Zunahme der Güte erwartet werden kann als ohne Berücksichtigung des Volumens. Im Gegensatz dazu erfolgt bei der Vermessung mit modellbasierten Messplänen die Teilung mit dem sample-normierten Kriterium, wodurch bei der Teilung die Anzahl an lokalen Messpunkten berücksichtigt wird. Ein Bereich mit schlechter Güte trotz vieler Messpunkte deutet auf eine hohe Nichtlinearität, die das aktuelle Modell nicht abbilden kann. Durch die bevorzugte Teilung in diesem Stellbereich kann dort die Nichtlinearität reduziert werden, wodurch eine stärkere Zunahme der Güte erwartet werden kann.

Weiterhin ist im Rahmen der Online-Vermessung nicht sichergestellt, dass die Messdaten in jeder Iteration gleichmäßig verteilt sind. Vielmehr erfolgt eine Stellbereichsanpassung bzw. werden neue Messdaten auch modellbasiert im Raum verteilt, wodurch eine Häufung von Messdaten in einem kleineren Bereich des globalen Modells vorkommen kann. Die Häufung der Messpunkte wird im Rahmen der Online-Methode genutzt, um das Verhältnis l_o des Teilungsalgorithmus zu bestimmen, indem dieses für jede Dimension $o = 1, \dots, R - 1$ an den Schwerpunkt (siehe z.B. [22]) der relevanten gewichteten Regressoren $s_{LMj,o}$ im zugehörigen lokalen Modell angepasst wird:

$$\begin{aligned} s_{LMj,o} &= \frac{\sum_{i=1}^N \Phi_j(\mathbf{z}_i) \cdot x_{o,i}}{\sum_{i=1}^N \Phi_j(\mathbf{z}_i)} \\ &= \frac{\Phi_j(\mathbf{z})^T \cdot \mathbf{x}_o}{\text{tr}(\mathbf{W}_j)} \end{aligned} \quad (5.65)$$

Hierbei ist $\Phi_j(\mathbf{z}) = [\Phi_j(\mathbf{z}_1), \dots, \Phi_j(\mathbf{z}_N)]^T$ der Vektor der Einzelgewichtungsfunktionen für alle Messpunkte $i = 1, \dots, N$ des lokalen Modells j .

Die adaptive Teilung folgt gemäß Bild 5.5 zu dem neuen Teilungsverhältnis

$$l_o^* = \frac{s_{LMRj,o} - x_{o,\min}}{x_{o,\max} - x_{o,\min}}. \quad (5.66)$$

Die Dimension o_{split} , in welcher die Teilung durchgeführt wird, wird bei ONLOPOMOT basierend auf den Seitenlängen der Hyperquader ermittelt:

$$o_{\text{split}} = \arg \max_{o \in 1, \dots, R-1} \Delta_{LMj,o} \quad (5.67)$$

Vorteile hierbei sind die schnellere Bestimmung, da nicht in alle Dimensionen geteilt werden muss, sowie die Vermeidung von schmalen lokalen Modellen, welche bei wenigen Messdaten ein schlechtes Extrapolationsverhalten aufweisen.

Netzmodellvereinigung

Neben der Teilung von lokalen Hyperquadern muss im Rahmen der Online-Vermessung auch die Vereinigung der lokalen Hyperquader berücksichtigt werden. Durch iterativ hinzukommen-

den Messdaten kann eine ursprünglich durchgeführte Teilung sich als nicht mehr relevant herausstellen. Der Vereinigungsalgorithmus für lokal polynomiale Modelle besteht dabei, angelehnt an den Vereinigungsalgorithmus von Junge [77] für lokal lineare Modelle, aus drei Anforderungen an die lokalen Modelle und Hyperquader:

1. Die zu vereinigenden lokalen Hyperquader müssen nebeneinander liegen
2. Die zu vereinigenden lokalen Hyperquader müssen vereinigt wieder ein Hyperquader ergeben
3. Die zu vereinigenden lokalen Modelle müssen ähnliche Parameter haben

Aus allen M vorhandenen Hyperquadern werden die zwei lokalen Hyperquader LMR_i und LMR_j (mit $i = 1, \dots, M$, $j = 1, \dots, M$ und $i \neq j$) bestimmt, für welche alle drei Bedingungen gegeben sind.

Zur Erfüllung der ersten Anforderung muss aus allen möglichen Dimensionen $o = 1, \dots, R - 1$ die Dimensionen m (also $m \in o$) bestimmt werden, bei denen der Abstand zwischen den beiden zu vereinigenden Mittelpunkten genau der halben Summe der beiden Seitenlängen in den Dimensionen m entspricht und gleichzeitig in allen anderen Dimensionen n (also $(n \in o) \wedge (n \notin m)$) die Mittelpunkte gleich sind:

$$|z_{0,\text{LMR}_i,m} - z_{0,\text{LMR}_j,m}| \stackrel{!}{=} \frac{1}{2} (\Delta_{\text{LMR}_i,m} + \Delta_{\text{LMR}_j,m}) \quad (5.68)$$

$$z_{0,\text{LMR}_i,n} \stackrel{!}{=} z_{0,\text{LMR}_j,n} \quad (5.69)$$

Hierbei sind z_{0,LMR_i} und z_{0,LMR_j} die beiden Mittelpunkte der zu vereinigenden Hyperquader sowie Δ_{LMR_i} und Δ_{LMR_j} die entsprechenden Seitenlängen. Durch die Sicherstellung, dass beide lokale Modelle nebeneinander liegen, wird verhindert, dass ein neues vereinigtes lokales Modell mit einem Mittelpunkt in einer nicht aktiven Region des lokalen Modells entsteht.

Die zweite Anforderung wird erfüllt, indem die Seitenlängen Δ_{LMR_i} und Δ_{LMR_j} der zu vereinigenden Hyperquader verglichen werden. Zwei vereinigte Hyperquader ergeben nur dann wieder ein Hyperquader, wenn in allen nicht vereinigten Dimensionen n die Seitenlängen paarweise gleich lang sind:

$$\Delta_{\text{LMR}_i,n} \stackrel{!}{=} \Delta_{\text{LMR}_j,n} \quad (5.70)$$

Einzig die Seitenlängen der zu vereinigenden Dimension m dürfen unterschiedlich lang sein. Existieren zwei Hyperquader i und j , für die alle Bedingungen der Gleichungen (5.68) - (5.70) erfüllt sind, können diese von ihrer Struktur her vereinigt werden, da sie vereinigt wieder ein Hyperquader ergeben.

Eine Vereinigung durchzuführen macht jedoch erst Sinn, wenn die jeweils in den lokalen Hyperquadern i und j gültigen lokalen Teilmodelle die dritte Anforderung erfüllen. Wären die beiden lokalen Teilmodelle stark verschieden, würde ein Vereinigen dieser die Nichtlinearität erhöhen, wodurch in einer der nächsten Iterationen wieder eine Teilung durchgeführt werden müsste. Eine Vereinigung ist somit nur sinnvoll, wenn die lokalen Teilmodelle ähnlich zueinander sind.

Die Bewertung der Ähnlichkeit der Parameter zweier Modelle kann mithilfe einer geometrischen Interpretation erfolgen. Ein lokales Teilmodell, welches linear in seinen Parametern ist, folgt zu

$$\hat{y}_{\text{LMR}i} = b_{0,\text{LMR}i} + b_{1,\text{LMR}i}x_1 + \dots + b_{R-1,\text{LMR}i}x_{R-1}. \quad (5.71)$$

Durch Umstellen kann das Modell in eine Hyperebene in Koordinatenform überführt werden. Eine Hyperebene besteht hierbei aus allen Punkten, deren Koordinaten $(\hat{y}_{\text{LMR}i}, x_1, \dots, x_{R-1})$ die folgende Gleichung erfüllen:

$$\hat{y}_{\text{LMR}i} - b_{1,\text{LMR}i}x_1 - \dots - b_{R-1,\text{LMR}i}x_{R-1} = b_{0,\text{LMR}i} \quad (5.72)$$

Durch Aufstellen des Normalenvektors

$$\mathbf{n}_{\text{LMR}i} = \begin{bmatrix} 1 \\ -b_{1,\text{LMR}i} \\ \vdots \\ -b_{R-1,\text{LMR}i} \end{bmatrix} \quad (5.73)$$

sowie Bestimmung des Abstands vom Koordinatenursprung anhand von $|b_{0,\text{LMR}i}| / |\mathbf{n}_{\text{LMR}i}|$ kann die Lage der Ebene eindeutig bestimmt werden. Hierbei wird deutlich, dass die Lage ausschließlich von den Parametern des Modells abhängt. Die Ähnlichkeit zweier Modelle kann dadurch anhand der Parallelität der zugehörigen Hyperebenen sowie dem Abstand derer vom Koordinatenursprung bewertet werden.

Zwei Hyperebenen $\text{LMR}i$ und $\text{LMR}j$ sind parallel, wenn der Winkel zwischen den Normalenvektoren $\angle \mathbf{n}_{\text{LMR}i}, \mathbf{n}_{\text{LMR}j} = 0^\circ$ ist. Dieser kann mithilfe des Skalarprodukts

$$\langle \mathbf{n}_{\text{LMR}i}, \mathbf{n}_{\text{LMR}j} \rangle = |\mathbf{n}_{\text{LMR}i}| |\mathbf{n}_{\text{LMR}j}| \cos(\angle \mathbf{n}_{\text{LMR}i}, \mathbf{n}_{\text{LMR}j}) \quad (5.74)$$

berechnet werden. Beträgt zusätzlich der Unterschied im Abstand der Hyperebenen vom Koordinatenursprung

$$\frac{|b_{0,\text{LMR}i}|}{|\mathbf{n}_{\text{LMR}i}|} - \frac{|b_{0,\text{LMR}j}|}{|\mathbf{n}_{\text{LMR}j}|} = 0, \quad (5.75)$$

dann sind beide Hyperebenen genau gleich.

Eine Ähnlichkeit der Hyperebenen, und somit der zugrundeliegenden lokalen Teilmodelle bzw. derer Parameter, wird durch Vorgabe von zulässigen Grenzwerten für den Unterschied in Winkel (Δ_{winkel}) und Offset (Δ_{offset}) bestimmt. Der aus den Normalenvektoren bestimmte Winkel darf den vorzugebenden Grenzwert Δ_{winkel} nicht überschreiten:

$$\arccos\left(\frac{\mathbf{n}_{\text{LMR}i}^T \mathbf{n}_{\text{LMR}j}}{|\mathbf{n}_{\text{LMR}i}| |\mathbf{n}_{\text{LMR}j}|}\right) \stackrel{!}{\leq} \Delta_{\text{winkel}}. \quad (5.76)$$

Wenn zusätzlich der Offset beider Hyperebenen sich nicht um mehr als die Schwelle Δ_{offset} unterscheidet

$$\left| \frac{|b_{0,\text{LMR}i}|}{|\mathbf{n}_{\text{LMR}i}|} - \frac{|b_{0,\text{LMR}j}|}{|\mathbf{n}_{\text{LMR}j}|} \right| \stackrel{!}{\leq} \Delta_{\text{offset}}, \quad (5.77)$$

dann sind beide Hyperebenen und somit auch beide lokale Modelle bzw. deren Parameter ähnlich.

Die lokalen Teilmodelle, welche paarweise alle Bedingungen erfüllen und somit sowohl von der Struktur als auch von den Parametern geeignet sind, können vereinigt werden. Existieren mehrere Paare, werden die zwei lokalen Teilmodelle vereinigt, welche die Bedingungen für Parameterähnlichkeit am besten erfüllen. Ausgehend von dem mit den vereinigten Hyperquadern entstehenden neuen Hyperquader können dann die Parameter des neuen Teilmodells im größeren Hyperquader neu bestimmt werden.

5.4.4 Online-Adaption der Ordnung der Modelleingangsgrößen

Die Modellkomplexität bei Netzmodellen ergibt sich neben der Modellteilung durch die Anzahl R an möglichen Regressoren x_i (siehe Gleichung (5.12)), welche sich aus der Ordnung S sowie Wechselwirkung der statischen Modelleingangsgrößen ergibt. Für die lokalen Modellausgangsgrößen folgt gemäß Gleichung (5.9):

$$\hat{y}_j(\mathbf{u}) = \begin{bmatrix} b_{0,j}, b_{1,j}, \dots, b_{R-1,j} \end{bmatrix} \cdot \begin{bmatrix} u_1^0 u_2^0 \dots u_{p-1}^0 u_p^0 \\ u_1^{k_1} u_2^{l_1} \dots u_{p-1}^{m_1} u_p^{n_1} \\ \vdots \\ u_1^{k_{R-1}} u_2^{l_{R-1}} \dots u_{p-1}^{m_{R-1}} u_p^{n_{R-1}} \end{bmatrix} \quad (5.78)$$

mit $(k_i + l_i + \dots + m_i + n_i) \leq S_j$ für $i = 1, \dots, R - 1$

ONLOPOMOT passt auch die sich aus den lokalen Regressoren ergebende Komplexität an. Jedes lokale Modell welches erstmalig trainiert wird, also sowohl das initiale Netzmodell mit nur einem lokalen Modell als auch jedes neu hinzukommende lokale Modell nach einer Teilung, wird mit einer statischen Ordnung von $S_j = 0$ trainiert, wodurch ein konstantes Modell $\hat{y}_j = b_{0,j}$ ermittelt wird.

Anschließend wird in jeder Iteration für alle lokalen Modelle die Notwendigkeit und Möglichkeit einer Erhöhung der maximal zulässigen Ordnung S_j geprüft. Eine Notwendigkeit liegt vor, wenn das lokale Modell die lokale Zielgüte $J_{\text{RMSE,lokal}}$ nicht erreicht. Daraufhin wird die Möglichkeit der Erhöhung geprüft. Diese ist gegeben, wenn die Anzahl an effektiven Datenpunkten im lokalen Modell größer als die Anzahl der Regressoren bei erhöhter Ordnung $S_j + 1$ ist. In jeder Iteration wird bei allen lokalen Teilmodellen, für die beide Bedingungen erfüllt sind, die Ordnung S_j um 1 erhöht und die Modellparameter neu bestimmt.

5.4.5 Gesamtstrategie zur Online-Modellkomplexitätsadaption

Bevor die Vermessung gestartet wird, muss die Vorgabe der Zielwerte der zu erreichenden Güten $G_{\text{RMSE,global,v}}$, $G_{\text{RMSE,global,t}}$, $G_{\text{RMSE,global-lokal,t}}$ und $G_{\text{RMSE,lokal,t}}$ erfolgen. Die Werte stellen die von der automatischen Vermessung zu erreichenden Güten dar und müssen im Voraus so gewählt werden, wie die gewünschte Modellgüte für den zugrundeliegenden Prozess sein soll. Ein Setzen aller Fehlerwerte zu 0 ist zwar prinzipiell möglich, führt jedoch dazu, dass in jeder Iteration alle Anpassungen durchgeführt werden, da ein Modellfehler von 0 in Realität nie auftreten wird. Da dies jedoch zu einer erhöhten Rechenzeit ohne Informationsgewinn führt, stellt die

Wahl der Zielwerte für die Modellfehler einen Trade-Off zwischen *so groß wie möglich* und *so klein wie nötig* dar.

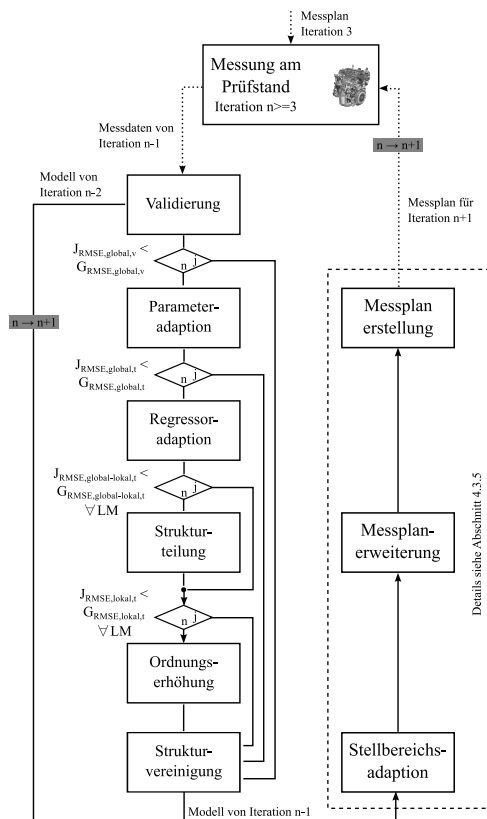


Bild 5.6: Gesamtstrategie zur Online-Identifikation von Modellen mit ONLOPOMOT. Die Modellstrukturadaption erfolgt kontinuierlich durch Integration in den iterativen Vermessungsprozess.

Ausgehend von den ersten Messdaten wird in der 2. Iteration eine initiale Identifikation durchgeführt und ein erstes Modell mit der stationären Regressorordnung $S = 0$ (und somit nur den Offset-Parameter b_0) und einem lokalen Modell ermittelt. Sobald neue Daten vermessen worden sind, startet die 3. Iteration. Ab dieser beginnt die in Bild 5.6 dargestellte Modellanpassungsstrategie, indem die folgenden Schritte iterativ ausgeführt werden:

1. Zuerst wird eine Validierung des alten Modells mit den neu hinzukommenden Messdaten durchgeführt, um den globalen Validierungsfehler $J_{\text{RMSE_global,v}}$ für diese Daten zu bestimmen.

2. Ist dieser kleiner als eine vorgegebene Grenze $G_{\text{RMSE,global,v}}$, dann ist das Modell zur Prozessdarstellung geeignet, und eine Erhöhung der Komplexität ist nicht notwendig. Vielmehr wird durch eine eventuelle Strukturvereinigung eine Reduktion der Komplexität versucht sobald mehr als ein Teilmodell vorhanden ist. Ist der Fehler größer als $G_{\text{RMSE,global,v}}$, erfolgt eine Parameteranpassung mit dem LASSO-Verfahren aus Abschnitt 5.4.1 unter Beibehaltung des Gewichtungsparameters λ aus Abschnitt 5.4.2.
3. Falls der globale Trainingsfehler $J_{\text{RMSE,global,t}}$ im Anschluss an die Parameteradaption klein genug ist, wird die Komplexitätsanpassung beendet und erneut eine Vereinigung der lokalen Teilmodelle versucht. Ist dagegen der Trainingsfehler immer noch zu hoch, wird eine Regressorauswahl mit dem LASSO-Verfahren durchgeführt, indem der Gewichtungsparameter λ neu bestimmt wird.
4. Erreicht die global-lokale Trainingsgüte $J_{\text{RMSE,global-lokal,t}}$ im Anschluss an die Regressoradaption den Zielwert, wird eine Strukturteilung nicht durchgeführt, sondern zur nächsten Adaption (Ordnungserhöhung) übergegangen. Wird dagegen die einzuhaltende Zielgüte immer noch nicht erreicht, wird das lokale Modell mit dem größten Fehler $J_{\text{RMSE,global-lokal,t}}$ entsprechend Abschnitt 5.4.3 geteilt.
5. Als letzte Möglichkeit der Strukturanpassung wird zusätzlich die maximal zulässige statische Ordnung S der Modelleingangsgrößen erhöht (siehe Abschnitt 5.4.4), falls der lokale Trainingsfehler $J_{\text{RMSE,lokal,t}}$ zu groß ist. Diese erfolgt für alle lokalen Teilmodelle mit genügend lokalen Messdaten. Die maximal mögliche Komplexitätserhöhung für diese Iteration ist erreicht und abschließend wird eine Vereinigung der lokalen Teilmodelle angestrebt, um die Komplexität weiter zu reduzieren.

Im Anschluss wird auf eine erneute Validierung mit neuen Messdaten gewartet. Nach Abschluss der Strukturadaption wird auf den Abschluss der parallel verlaufenden Vermessung gewartet. Sobald neue Messdaten vorliegen ($n \rightarrow n + 1$), erfolgt der Ablauf der Strukturadaption von Beginn an. Mithilfe der beschriebenen Methodik wird das Modell bei Bedarf iterativ an den Prozess angepasst, indem die Struktur kontinuierlich adaptiert wird. Dadurch kann eine maximale Verbesserung des Modells im Hinblick auf sowohl Bias als auch Varianz erreicht werden. Durch die laufende Erhöhung bzw. bei Bedarf auch die Verringerung der globalen Modellkomplexität wird diese zielgerichtet angepasst. Wegen der Integration der Validierung in die iterative Adaption wird eine zusätzliche Validierung im Nachgang mit separaten Messdaten überflüssig, da bereits in jeder Iteration eine Validierung mit neuen Messdaten durchgeführt wird.

5.4.6 Beispiel der Online-Identifikation

Zum Vergleich der Offline-Identifikation mit dem neu entwickelten Online-Verfahren ONLO-POMOT wird im Folgenden erneut die statische Peaks-Funktion aus Abschnitt 4.3.6 als zu identifizierender Prozess simuliert. Vermessen werden $Z = 79$ Iterationen, wobei die erste aus 9 Messpunkten der initialen Variationsraumvermessung und alle folgenden aus iterativ 5 neuen Messpunkten nach der DTS-Versuchsplanung besteht. Als Zielwerte für den zulässigen

Modellfehler werden $J_{\text{RMSE,lokal,t}} \leq 0.01$, $J_{\text{RMSE,global-lokal,t}} \leq 0.008$ und $J_{\text{RMSE,global,t}} \leq 0.02$ vorgegeben.

Neben der Online-Modellierung mit ONLOPOMOT erfolgt zusätzlich in jeder Iteration die Offline-Bestimmung eines Modells mit dem LOPOMOT-Verfahren, um einen iterativen Vergleich beider Verfahren durchführen zu können. Dem LOPOMOT-Modell wird dabei als maximale Modellordnung und maximale Anzahl an lokalen Modellen die jeweilige Größe bzw. Anzahl des ONLOPOMOT-Modells vorgegeben, ohne jedoch die Teilungsstruktur bzw. die Regressoren vorzugeben. Dadurch kann das LOPOMOT-Modell die Teilungsstruktur (bis zur maximalen Anzahl an lokalen Modellen) sowie die Auswahl der Regressoren (bis zur maximalen Anzahl an Regressoren) durchführen. Durch die Identifikation beider Modelle mit jeweils dem selben Datensatz kann der Einfluss des Identifikationsverfahrens unabhängig von dem zugrunde gelegten Messplan bewertet werden. Die Modelle werden mit einem zweiten unabhängigen Datensatz validiert, welcher aus 10000 Messpunkten eines LHS-Messplans besteht.

In Bild 5.7 sind die sich nach der letzten Iteration ergebenden a) Modellausgangsgrößen als *measured vs. predicted* Plot für das Training sowie die Validierung mit dem separaten Datensatz dargestellt. Weiterhin sind in b) die resultierenden Netzmodellteilungen mit den jeweiligen Höhenlinien des zugrundeliegenden Prozesses beider Verfahren dargestellt. Beide Verfahren können prinzipiell den Prozess modellieren.

Zum genaueren Vergleich der beiden Verfahren wird zusätzlich der iterative Verlauf der beiden Trainings- und Validierungsgüten in Bild 5.8 a) sowie der Verlauf der iterativen und kumulierten Zeitdauer zur Modellbildung in b) dargestellt. Das Online-Verfahren hat sowohl im Training als auch in der Validierung einen leicht kleineren Fehler als das Offline-Verfahren (siehe Tabelle 5.2), wobei das Offline-Verfahren die Zielgüte früher erreicht. Der leichte Unterschied in der Güte ist darauf zurückzuführen, dass das Online-Verfahren bei der Adaption sowohl die globale, als auch die global-lokale und lokale Güte bewertet. LOPOMOT berücksichtigt dagegen ausschließlich die globale Güte und adaptiert nicht weiter wenn deren Zielwert erreicht ist. Das Offline-Verfahren erreicht die Zielgüte früher als das Online-Verfahren, weil es jedesmal neu initiiert wird und eine für die zugrundeliegenden Messdaten optimale Modellstruktur auswählen kann. Das Online-Verfahren dagegen baut iterativ die Modellstruktur auf und muss dadurch von der Modellstruktur der vorhergehenden Iteration ausgehen. Hierdurch kann nicht sofort die optimale Modellstruktur bestimmt werden, sondern erst im weiteren Verlauf.

Tabelle 5.2: Vergleich der erreichten Güten für das ONLOPOMOT- und LOPOMOT-Verfahren.

	$J_{\text{RMSE,global,t}}$	$J_{\text{RMSE,global,v}}$
LOPOMOT	0.0193	0.0196
ONLOPOMOT	0.0145	0.0161

Im Gegensatz zu den Güten ist der Unterschied in der Modellierungsdauer signifikant:

$$\begin{aligned} t_{\text{LOPOMOT}} &= 2178.9 \text{ s} \\ t_{\text{ONLOPOMOT}} &= 263.1 \text{ s.} \end{aligned} \tag{5.79}$$

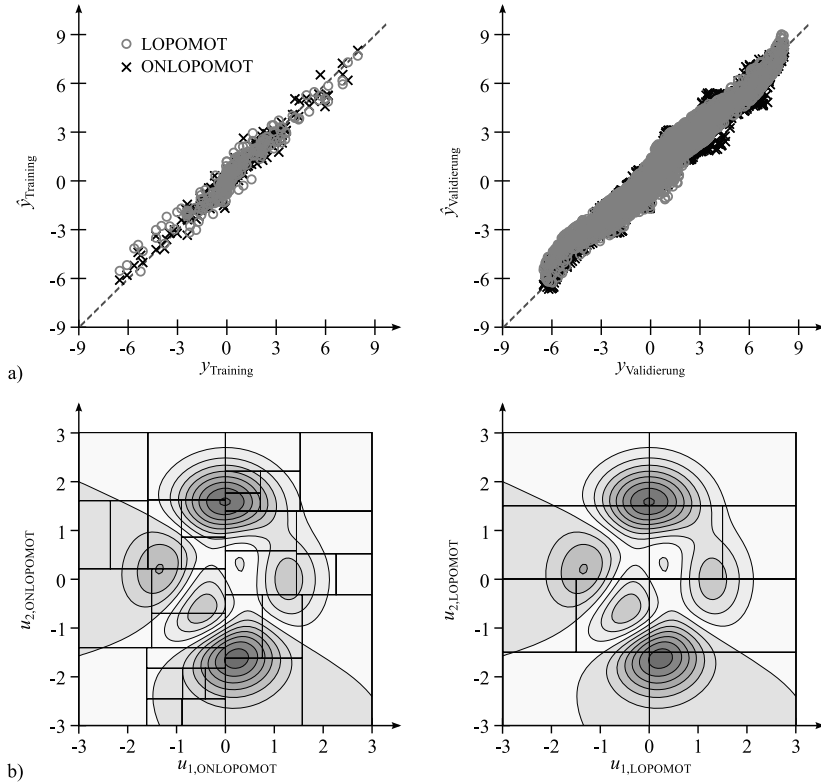


Bild 5.7: Vergleich des Ergebnisses der Modellbildung mit einem LOPOMOT- und ONLOPOMOT-Modell zum Ende der iterativen Vermessung. Die identifizierten Modelle weisen eine hohe Güte auf, sowohl im Training als auch in der Validierung (a). Mit ONLOPOMOT erfolgt eine gezieltere Anpassung der Teilungsstruktur an den zugrundeliegenden Prozess im Vergleich zu LOPOMOT (b).

ONLOPOMOT führt eine Modellbildung nur durch, wenn sie für die neuen Messpunkte nötig ist und setzt die Strukturbildung ausgehend von dem letzten Modell fort. Im Gegensatz dazu nutzt LOPOMOT immer alle Messpunkte, um eine komplett neue Identifikation sowie Modellstrukturbestimmung durchzuführen, wodurch die Dauer der Modellierung stetig zunimmt. In diesem Beispiel hat LOPOMOT ab der 18. Iteration in Summe länger für die Modellbildung benötigt als ONLOPOMOT.

Insgesamt wird deutlich, dass das neu entwickelte Verfahren zur iterativen Modellbildung von Prozessen gut geeignet ist. Durch den Einsatz der ONLOPOMOT-Modellierung kann die benötigte Zeit zur Modellbildung, auf Kosten einer nur gering schlechteren Modellgüte, signifikant reduziert werden. Dadurch kann die in Abschnitt 5.4 beschriebene Forderung, dass die Modell-

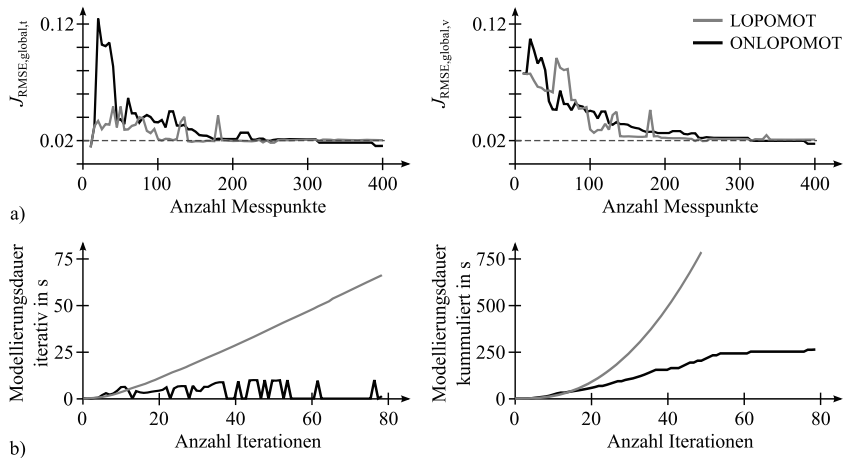


Bild 5.8: Vergleich der iterativen Güten und Modellierungsdauern, simuliert mit einem Rechner mit 2.40 GHz Prozessor in MATLAB R2011b. Beide Modellierungsarten erreichen die Zielgüte (a). Die Modellierungsdauer steigt bei LOPOMOT exponentiell, während ONLOPOMOT eine Anpassung des Modells nur bei Bedarf vornimmt (b).

bildung nicht länger dauern darf als die Messung, für einen wesentlich größeren Messdatensatz und somit für wesentlich mehr Prozessinformationen erfüllt werden.

5.5 Zusammenfassung

Die Wahl eines geeigneten Verfahrens zu Modellbildung eines Prozesses ist ausschlaggebend für die erreichbare Güte des Modells. Für die Anwendung im Rahmen einer Online-Vermessung müssen geeignete Erweiterungen vorgenommen werden, um die Forderung nach einer kurzen Modellierungsdauer zu erfüllen. Mit dem in diesem Kapitel vorgestellten ONLOPOMOT-Verfahren kann eine kontinuierlich adaptive Modellbildung mit dem iterativ größer werdenden Trainingsdatensatz erfolgen, bei der neben den Modellparametern auch die Modellstruktur iterativ angepasst wird.

ONLOPOMOT basiert auf der Modellbildung mit lokal polynomialen Modellen (LOPOMOT). Lokale Teilmodelle werden zur Strukturbestimmung mit globalen, global-lokalen sowie lokalen Gütekriterien zur Modellanalyse bewertet. Das Netzmodell wird dabei basierend auf Kriterien, welche aufgrund der iterativen Vergrößerung des Messdatensatzes zusätzlich das Volumen der Hyperquader bzw. die Anzahl der lokalen Messpunkte implizieren, geteilt. Durch die Erweiterung kann die globale Modellgüte für zukünftige Iterationen zielgerichtet verbessert werden. Die Parameter des Modells werden iterativ mit dem RLS-Verfahren für alle aktiven Teilmodelle angepasst, wodurch ein Vergessen der Teilmodelleigenschaften vermieden wird. Die Modellkomplexität wird iterativ durch die Regressoradaption mit dem LASSO-Verfahren, die Struk-

turanpassung der Netzmodellteilung sowie Erhöhung der zulässigen maximalen Regressorordnung angepasst.

Durch die Kombination einer zunehmenden Datensatzgröße mit der zielgerichteten Anpassung der Modellkomplexität des ONLOPOMOT-Verfahrens werden sowohl der Bias- als auch der Varianzfehler reduziert. Hierdurch löst sich das Bias-Varianz-Dilemma der Offline-Verfahren auf, und der Gesamtfehler des Modells wird kontinuierlich mit zunehmender Datensatzgröße kleiner.

6 Anwendungsbeispiel der Online-Vermessung

Die Einbettung der in den vorhergehenden Kapiteln entwickelten Methoden in eine durchgehende Gesamtstrategie kann zur zielgerichteten und schnellen Online-Vermessung genutzt werden. Alle im Rahmen dieser Arbeit entwickelten Algorithmen wurden in ein Software-Paket integriert und nicht nur aus theoretischen Gesichtspunkten untersucht, sondern auch im praktischen Einsatz validiert und stehen zur Weiterverwendung bereit. Dies wird im Folgenden anhand eines Beispiels der lokalen statischen Online-Modellbildung verdeutlicht.

Die im Rahmen der zwei FVV-Projekte „Online-Methoden“ und „Online-Methoden II“ entwickelte Online-Methodik ermöglicht es, einen signifikanten Teil des Vermessungs- und Modellierungsprozesses automatisiert direkt am Prüfstand durchzuführen.

Die Kombination sowohl der DoE- als auch der Prüfstandsautomatisierungs-Software in eine Gesamtmethodik bewirken die in den vorherigen Kapiteln beschriebenen Verbesserungen gegenüber dem herkömmlichen Vorgehen. Durch die Integration dieser kombinierten Gesamtmethodik am Motorenprüfstand läuft der gesamte Prozess der Motorvermessung und Modellbildung automatisch ab.

Verschiedene praktische Beispiele, welche im Rahmen der Verifikation der entwickelten Gesamtmethodik am institutseigenen Motorenprüfstand (siehe Anhang A) entstanden sind, können den beiden Abschlussberichten [92, 93] entnommen werden.

An dieser Stelle werden wesentliche Aspekte der Online-Methodik anhand eines Beispiels der lokalen statischen Modellbildung veranschaulicht. In Bild 6.1 ist vereinfacht das Vorgehen für a) das Online-Vermessen sowie b) die im Folgenden vorgestellte Validierung dargestellt. Ausgehend von Messdaten einer Rastervermessung wird im ersten Schritt ein detailliertes LOPOMOT-Offline-Modell des Verbrennungsmotors (Details zum eingesetzten Motor siehe Anhang A) identifiziert [148, 113]. Das so entstandene Offline-Modell des Verbrennungsmotors wird zur Validierung der entwickelten Methodik genutzt, indem die Ausgangsgrößen für die iterativen Messpunkte im Verlauf der Online-Vermessung generiert werden. Vorteil einer solchen Validierung ist, dass die Offline- und Online-Strategie zueinander detailliert bewertet werden können, ohne dabei durch den Prüfstand (z.B. andere Konditionierung des Kraftstoffs) oder den Versuchsträger (z.B. andere Beladung des Partikelfilters) ungewollt beeinflusst zu werden.

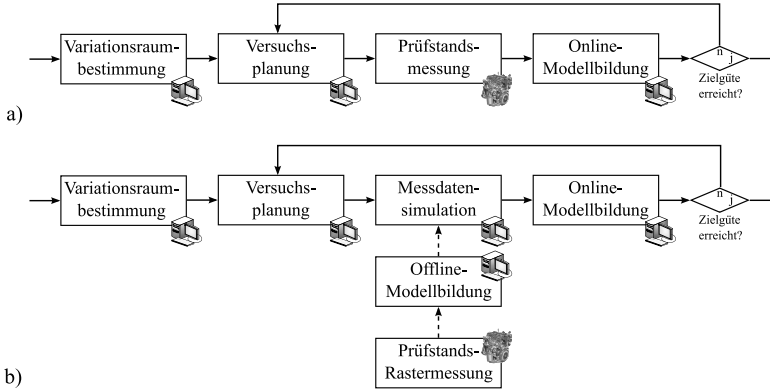


Bild 6.1: a) Vereinfacht dargestelltes Vorgehen der Online-Vermessung. b) Vorgehen zur Validierung der Methodik unabhängig von ungewollten Prüfstands- und Versuchsmotoreffekten in der Bewertung, wie z.B. Partikelfilterbelastung. Hierbei wird aus einer Prüfstands-Rastermessung im ersten Schritt ein Offline-Modell bestimmt, mit welchem im Anschluss während einer iterativen Online-Modellbildung immer wieder neue Messdaten generiert werden können.

6.1 Parametrierung der Online-Vermessung

Vermessen werden die drei Ausgangsgrößen \dot{m}_{NO_x} -Massenstrom \dot{m}_{NO_x} , Ruß-Massenstrom \dot{m}_{RuB} und das Motormoment M_{Mot} in dem lokalen Arbeitspunkt ($2250 \text{ min}^{-1} / 18 \frac{\text{mm}^3}{\text{Hub}}$). Als Stell- sowie Modelleingangsgrößen werden die Luftmasse m_{Luft} , der Ladedruck p_2 , der Einspritzwinkel der Haupteinspritzung φ_{HE} sowie die Temperatur am Einlass T_2 verwendet.

Als Grenzgrößen des Variationsraums werden für die Emissionen NO_x und Ruß Grenzen vorgegeben ($\dot{m}_{\text{NO}_x} < 30 \text{ mg/s}$, $\dot{m}_{\text{RuB}} < 0.08 \text{ mg/s}$). Dadurch kann bereits bei der Online-Vermessung sichergestellt werden, dass das optimierte Kennfeld sich nur über Stellbereiche aufspannt, welche Emissionen aufweisen die kleiner sind als die vorgegebenen Grenzwerte.

Um einen Einsatz des Modells im Rahmen einer Sollwertkennfeldoptimierung zu ermöglichen, wird bei der Versuchsplanung das global optimierte lokale Pareto-Optimum gemäß der Zielfunktion aus Gleichung (4.52) berücksichtigt. Die Optimierung erfolgt dabei mit dem NSGA-II Algorithmus und bestimmt eine Pareto-Front bestehend aus 10 Punkten, wobei innerhalb einer Iteration maximal zwei Optimierungsschritte (zwei neue Generationen) durchgeführt werden. In der folgenden Iteration werden nur die Stellgrößen des Optimums als MO-Messplanpunkt vermessen, welche am weitesten entfernt von allen bis dahin vermessenen Messpunkten liegen. Dadurch kann sichergestellt werden, dass die vermessene Pareto-Front iterativ verbessert wird.

Die Vermessung erfolgt gemäß Abschnitt 4.3.5 und richtet sich nach der sich ändernden Modellstruktur der drei Modelle. Entweder wird ein DTS- und für jede Ausgangsgröße ein MQD-Messplanpunkt vermessen oder ein MO- und für jede Ausgangsgröße je ein MPC- und ein

MQI-Messplanpunkt. Je nach Modelländerung der drei Ausgangsgrößen werden dadurch iterativ 4 – 7 neuen Messpunkte vermessen.

Während der iterativen Vermessung wird laufend die Modellkomplexität entsprechend der in Abschnitt 5.4.5 beschriebenen Strategie adaptiert. Als Grenze wird für die einzelnen Ausgangsgrößen

$$\begin{aligned} G_{M_{\text{Mot}}, \text{RMSE}} &= 0.21 \text{ Nm} \\ G_{\dot{m}_{\text{NO}_x}, \text{RMSE}} &= 1.04 \text{ mg/s} \\ G_{\dot{m}_{\text{RuB}}, \text{RMSE}} &= 0.01 \text{ mg/s} \end{aligned} \quad (6.1)$$

vorgegeben, welche normiert auf den Einheitsbereich der jeweiligen Ausgangsgröße (gemäß Anhang F) in allen Fällen 3 % entsprechen.

6.2 Verlauf der Online-Vermessung

Die Vermessung erfolgt mit 80 Iterationen als Abbruchkriterium, um neben dem schnellen Erreichen der Zielgüte auch den Einfluss von weiteren Messungen beurteilen zu können. In Bild 6.2 a) sind der gesamte Stellgrößenverlauf sowie in b) die Verläufe der Modellstrukturänderung für alle drei Modellausgangsgrößen dargestellt.

In der ersten Iteration wird der Variationsraum mit einem Hüllraster-Messplan mit 3 Stufen für die 4 Stellgrößen bestimmt, woraus sich die ersten 81 Messpunkte ergeben. Die sich aus den vorgegebenen Emissionsgrenzen ($\dot{m}_{\text{NO}_x} < 30 \text{ mg/s}$, $\dot{m}_{\text{RuB}} < 0,08 \text{ mg/s}$) ergebenden zulässigen Stellgrößenbereiche folgen für diesen Arbeitspunkt zu:

$$\begin{aligned} 0.33 \frac{\text{g}}{\text{Hub}} &\leq m_{\text{Luft}} \leq 0.62 \frac{\text{g}}{\text{Hub}} \\ 1.08 \text{ bar} &\leq p_2 \leq 1.36 \text{ bar} \\ 1.63 \text{ °KWvOT} &\leq \varphi_{\text{HE}} \leq 9.28 \text{ °KWvOT} \\ 37.07 \text{ °C} &\leq T_2 \leq 94.26 \text{ °C} \end{aligned} \quad (6.2)$$

Ausgehend von den Messdaten des Variationsraums wird ein erster raumfüllender Messplan mit 3 Messpunkten erstellt und vermessen, während parallel das erste Modell, welches sich aus den ersten 81 Messpunkten ergibt, identifiziert wird. Anschließend wird mit iterativen Online-Messplänen vermessen, wodurch in jeder Iteration 4 – 7 neue Messpunkte hinzukommen.

Es wird deutlich, dass ab der 49., 29. bzw. 27. Iteration nur noch wenige signifikante Modelländerungen bei den drei Ausgangsgrößen erfolgen. Die Modelle haben qualitativ ihre Zielgüte erreicht und im Folgenden werden überwiegend nur noch Messpläne vermessen, welche für die zugrundeliegenden Modellstrukturen eine gute Messpunktverteilung sowie hohe Güte der Pareto-Front erreichen (Messpläne nach den MQI-, MPC- sowie MO-Verfahren). Die Parameteradaption wurde für die drei modellierten Ausgangsgrößen mit 60, 32 bzw. 32 mal am Häufigsten durchgeführt. Dies ist auf die Feinanpassung zurückzuführen, welche auch im späteren Verlauf der Vermessung bei bereits fester Modellstruktur erfolgt. Im Vergleich dazu wurde eine

Regressoradaption 41, 18 bzw. 10 mal, eine Modellteilung 19, 15 bzw. 10 mal und eine Erhöhung der statischen Ordnung 27, 15 bzw. 8 mal durchgeführt. 20, 48, bzw. 48-mal wurde aufgrund der hohen Validierungsgüte keine Modelladaption vorgenommen.

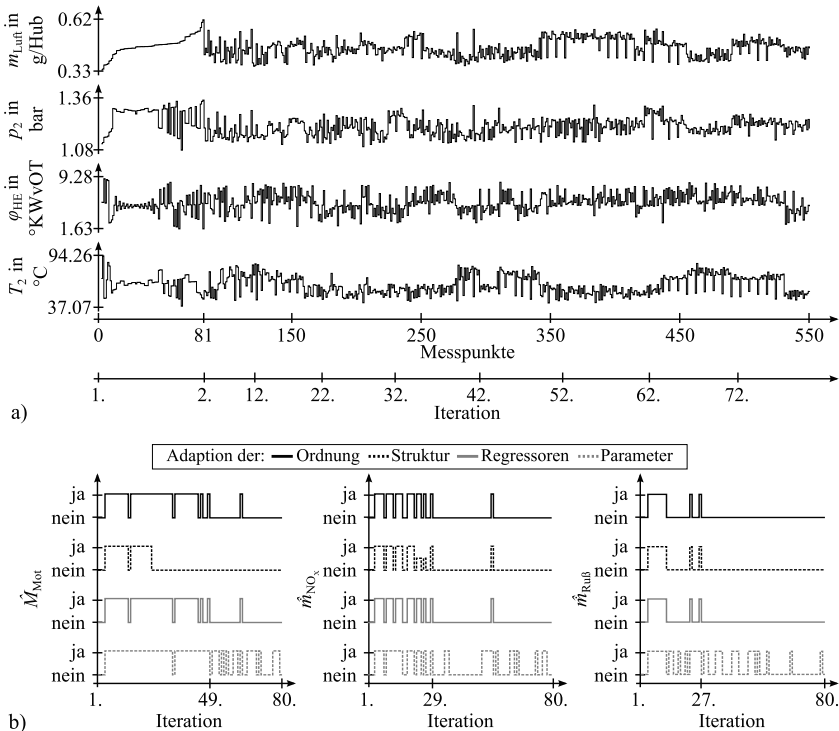


Bild 6.2: a) Stellgrößen der iterativen Vermessung für den lokalen Arbeitspunkt ($2250 \text{ min}^{-1} / 18 \frac{\text{mm}^3}{\text{Hub}}$) sowie b) Verlauf der Art der iterativen Modelladaption. Ab der 49., 29. bzw. 27. Iteration erfolgt die Vermessung überwiegend mit dem Ziel, ein gutes Optimum und eine gute Messpunktverteilung für die zugrundeliegenden Modellstrukturen zu erreichen.

6.3 Auswertung der Online-Vermessung

Bild 6.3 zeigt die resultierenden *measured vs. predicted* Plots der modellierten Ausgangsgrößen. In a) ist jeweils der Vergleich der mit dem Offline-Modell generierten Ausgangsgrößen zu den Ausgangsgrößen des Online-Modells (Training) und in b) der Vergleich der gemessenen rasterförmigen Ausgangsgrößen zu den Ausgangsgrößen des Online-Modells (Validierung)

dargestellt. Die Güten erreichen sowohl in dem Training als auch in der Validierung die vorgegebenen Zielwerte (siehe Tabelle 6.1).

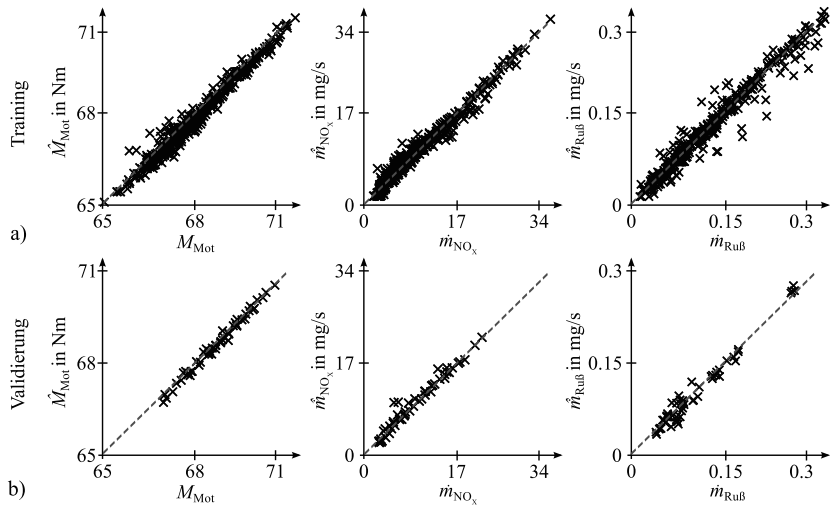


Bild 6.3: Darstellung der Güte des bestimmten Modells im Arbeitspunkt ($2250 \text{ min}^{-1} / 18 \frac{\text{mm}^3}{\text{Hub}}$) als measured vs. predicted Plots für a) das Training (Ausgangsgrößen der Trainingsdaten vs. Ausgangsgrößen des Online-Modells) sowie b) die Validierung (Ausgangsgrößen der Rastermessung vs. Ausgangsgrößen des Online-Modells). Das Modell weist sowohl im Training als auch in der Validierung eine hohe Güte auf.

Tabelle 6.1: Vergleich der erreichten J_{RMSE} -Güten für die Online-Vermessung der drei Ausgangsgrößen \dot{m}_{NO_x} , \dot{m}_{RuB} und M_{Mot} in dem lokalen Arbeitspunkt ($2250 \text{ min}^{-1} / 18 \frac{\text{mm}^3}{\text{Hub}}$).

	$J_{\text{RMSE,global,t}}$	$J_{\text{RMSE,global,v}}$
M_{Mot}	0.195 Nm bzw. 2.8 %	0.201 Nm bzw. 2.9 %
\dot{m}_{NO_x}	0.774 mg/s bzw. 2.2 %	0.913 mg/s bzw. 2.6 %
\dot{m}_{RuB}	0.009 mg/s bzw. 2.6 %	0.010 mg/s bzw. 2.8 %

Die nach Abschluss der Vermessung für den lokalen Arbeitspunkt ($2250 \text{ min}^{-1} / 18 \frac{\text{mm}^3}{\text{Hub}}$) identifizierten Modelle bestehen aus 20, 16 bzw. 11 lokalen Teilmodellen. Die maximalen statischen Ordnungen gemäß Gleichung 5.18 sind $S_{M_{\text{Mot}}} = 3$, $S_{\dot{m}_{\text{NO}_x}} = 2$ und $S_{\dot{m}_{\text{RuB}}} = 2$, wobei die lokalen Modelle jeweils eine unterschiedliche Anzahl an Regressoren haben. In den Bildern 6.4-6.6 sind die identifizierten Modelle der Ausgangsgrößen sowie die jeweils zugehörigen gemessenen Rastermesspunkte als 4-dimensionale Kennfelder dargestellt.

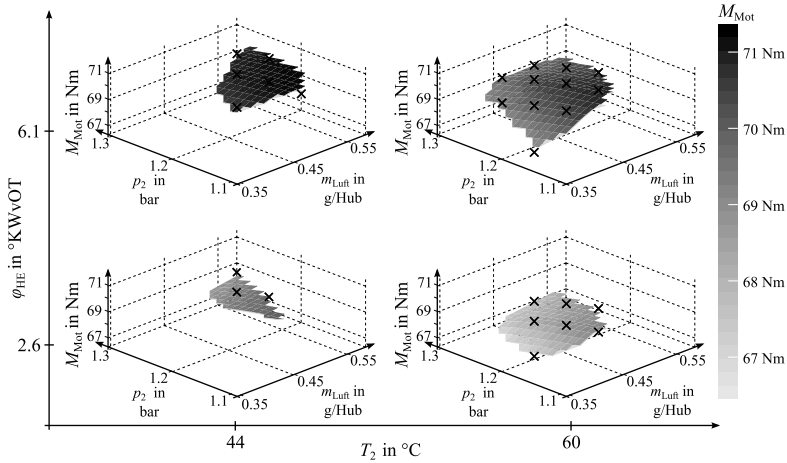


Bild 6.4: Verlauf des Drehmoments in Abhängigkeit der Stellgrößen Luftmasse m_{Luft} , Ladedruck p_2 , Einspritzwinkel der Haupteinspritzung φ_{HE} sowie Temperatur am Einlass T_2 für Messdaten einer Rastervermessung (Kreuze) im Arbeitspunkt ($2250 \text{ min}^{-1} / 18 \frac{\text{mm}^3}{\text{Hub}}$). Zusätzlich ist das mit der Online-Methodik identifizierte Modell dargestellt (Fläche).

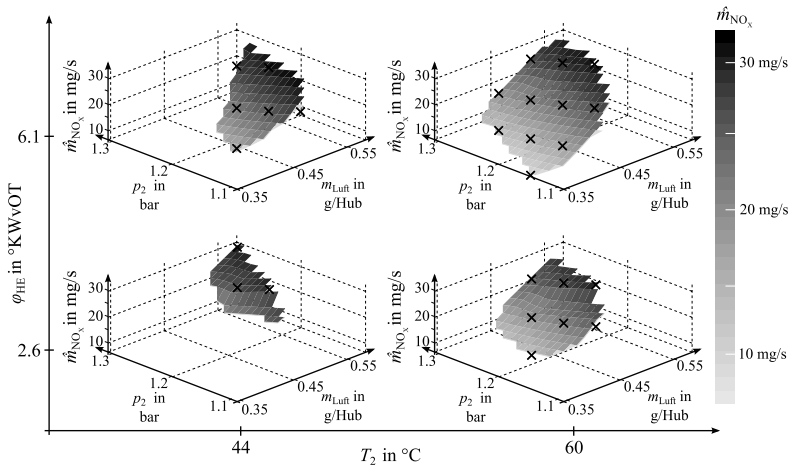


Bild 6.5: Verlauf des NO_x -Massenstroms in Abhängigkeit der Stellgrößen Luftmasse m_{Luft} , Ladedruck p_2 , Einspritzwinkel der Haupteinspritzung φ_{HE} sowie Temperatur am Einlass T_2 für Messdaten einer Rastervermessung (Kreuze) im Arbeitspunkt ($2250 \text{ min}^{-1} / 18 \frac{\text{mm}^3}{\text{Hub}}$). Zusätzlich ist das mit der Online-Methodik identifizierte Modell dargestellt (Fläche).

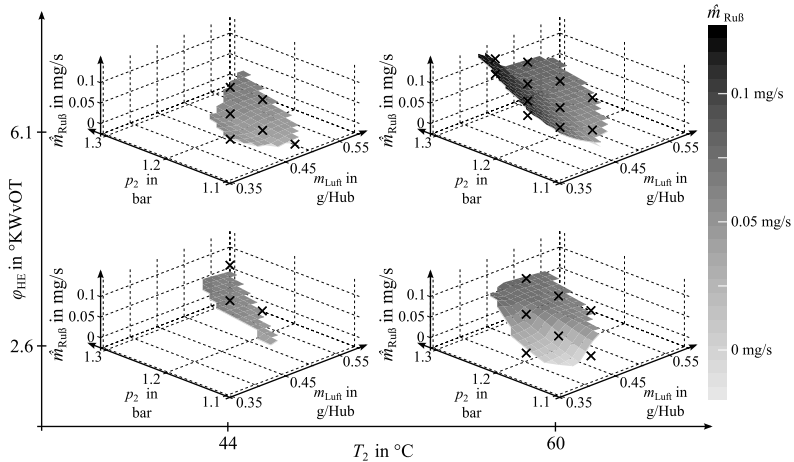


Bild 6.6: Verlauf des Ruß-Massenstroms in Abhängigkeit der Stellgrößen Luftmasse m_{Luft} , Ladedruck p_2 , Einspritzwinkel der Haupteinspritzung φ_{HE} sowie Temperatur am Einlass T_2 für Messdaten einer Rastervermessung (Kreuze) im Arbeitspunkt ($2250 \text{ min}^{-1} / 18 \frac{\text{mm}^3}{\text{Hub}}$). Zusätzlich ist das mit der Online-Methodik identifizierte Modell dargestellt (Fläche).

Im Gegensatz zum klassischen Vermessen weist die Online-Vermessung keinen wieder ansteigenden Validierungsfehler auf (siehe Bild 5.4 c)). Dies wird am Verlauf der Trainings- und Validierungsgüten der Modelle für die verschiedenen Iterationen bzw. an der dazugehörigen Anzahl an Messpunkten deutlich. Wie aus Bild 6.7 ersichtlich, weisen sowohl der Trainings- als auch der Validierungsfehler der drei Modelle iterativ insgesamt überwiegend eine abnehmende Tendenz auf. Hierdurch wird der abnehmende Bias- und Varianz-Fehler deutlich. Durch das iterative Online-Verfahren werden die Modelle zielgerichtet adaptiert und die Modellgüte maximiert. Die vereinzelte Zunahme des Trainings- bzw. Validierungsfehlers ist auf die Vermessungsstrategie zurückzuführen, welche bei Erreichen der Zielgüte keine Adaption des Modells mehr vornimmt aber trotzdem neue Messpunkte in Bereiche mit geringer Modellgüte erstellt.

Zusätzlich ist in Bild 6.7 auch die Güte eines LOPOMOT-Modells dargestellt, welches mit den gleichen iterativen Messdaten jedes mal neu bestimmt wird. Es wird deutlich, dass beide Verfahren geeignet sind, die Zielgüte zu erreichen, jedoch das Offline-Verfahren bei einer Neubestimmung des Modells in jeder Iteration eine bessere Güte aufweist. Dies ist auf zwei Eigenschaften der Online-Modellbildung zurückzuführen. Zum einen bestimmt das Online-Verfahren das Modell unter Vorgabe einer zu erreichenden Zielgüte. Sobald diese erreicht wird, erfolgt keine weitere Adaption. Dagegen bestimmt ein Offline-Modell immer für den vorliegenden Messdatensatz das beste Modell. Zum Zweiten wird die Modellkomplexität bei dem Online-Modell ausgehend von der Komplexität der vorherigen Iteration rekursiv adaptiert, während ein Offline-Modell immer neu initiiert werden muss und dadurch die für die vorliegenden Messdaten optimale Modellkomplexität neu bestimmt werden kann. Ein Online-Modell kann somit

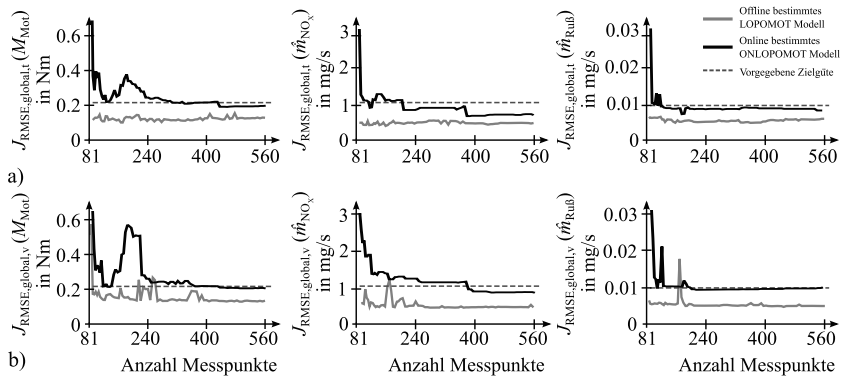


Bild 6.7: Verlauf der Trainings- und Validierungsgüten während der iterativen Online-Vermessung. Die verschiedenen Gütemaße sind in Gleichungen (5.55) - (5.59) beschrieben.

im besten Fall nur die selbe Güte erreichen, die ein Offline bestimmtes Modell aufweist. Der schlechteren Modellgüte (wobei diese besser als der vorgegebene Zielgütwert ist) steht der Vorteil gegenüber, eine signifikant schnellere Modellbildung direkt am Prüfstand während der laufenden Messung durchführen zu können, siehe Abschnitt 6.4.

Durch die in die iterative Vermessung integrierte Optimierung ergibt sich eine 3-dimensionale Pareto-Fläche. Diese ist in Bild 6.8 für a) die 27. und b) die letzte Iteration dargestellt (Punkte, in Grautönen skaliert über den spezifischen Kraftstoffverbrauch). Zusätzlich ist zwecks Validierung jeweils die im Nachgang mit dem globalen Optimierer bestimmte Pareto-Fläche eingezeichnet (Kreuze, in Grautönen skaliert über den spezifischen Kraftstoffverbrauch). In Bild 6.8 c) und d) sind die sich daraus für einen spezifischen Kraftstoffverbrauch von 250 g/kWh resultierenden Pareto-Fronten der spezifischen Emissionen dargestellt, auch hier wieder in c) für die 27. und d) die letzte Iteration sowie jeweils die im Nachgang offline bestimmte Pareto-Front. Aus den Darstellungen wird deutlich, dass mit der in die Online-Vermessung integrierten Optimierung die resultierende Pareto-Front sehr genau bestimmt werden kann. Diese Pareto-Fronten zeigen für den gewählten Arbeitspunkt den typischen Zusammenhang zwischen den NO_x - und Ruß-Emissionen: Die eine Emission kann nur auf Kosten der anderen minimiert werden.

Die bessere Bestimmung der Optima im Verlauf der iterativen Vermessung kann quantitativ bewertet werden, indem zu jedem Punkt der tatsächlichen Pareto-Fläche der Abstand (euklidische Norm) des am nächsten liegenden vermessenen Punktes bestimmt wird. Die Summe aller minimalen Abstände ist ein Maß für die Güte der vermessenen Pareto-Fläche, wobei je kleiner die Summe ist, desto besser ist diese vermessen. Hierbei ist zu beachten, dass die Abstände mit den auf den Einheitsbereich normierten Größen bestimmt werden (siehe Anhang F). Die bestimmten Werte sind dadurch zwar Einheitslos, können aber zu einer qualitativen Bewertung zueinander genutzt werden.

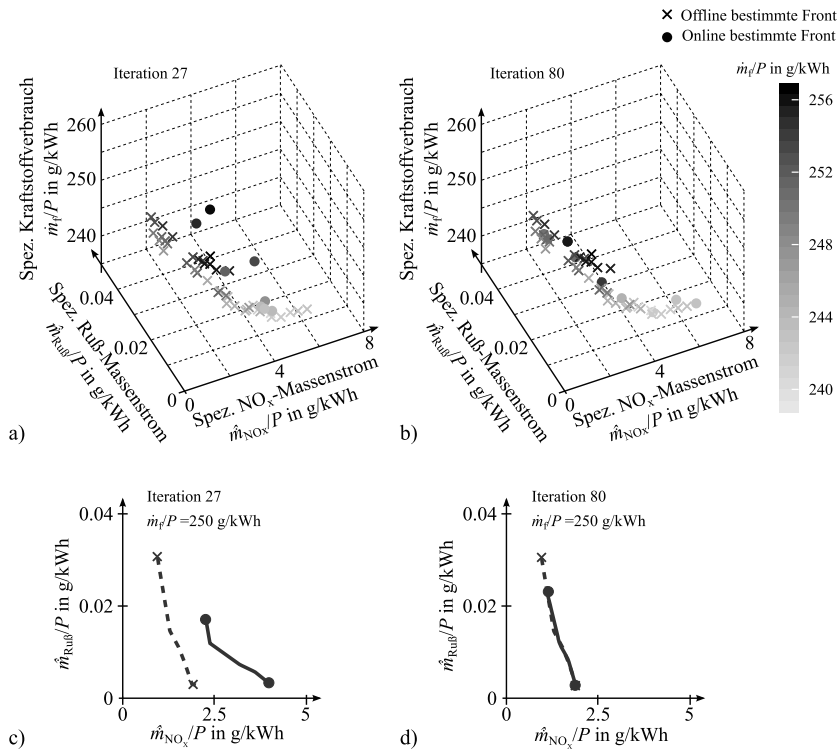


Bild 6.8: Resultierende 3-dimensionale Pareto-Fläche (Punkte) für die Optimierung der lokalen Zielfunktion aus 4.52 im Arbeitspunkt ($2250 \text{ min}^{-1} / 18 \frac{\text{mm}^3}{\text{Hub}}$) mit dem globalen NSGA-II Optimierer für a) 27. und b) die letzte Iteration der Online-Vermessung. Zusätzlich ist mit Kreuzen die im Nachgang offline bestimmte Fläche dargestellt. Die Grautöne skalieren mit dem jeweiligen spezifischen Kraftstoffverbrauch. Die genauere Bestimmung der Fläche wird in c) und d) deutlich. Hier ist die 3-dimensionale Pareto-Fläche auf eine 2-dimensionale Pareto-Front reduziert worden, indem der Schnitt durch die Fläche bei einem spezifischen Kraftstoffverbrauch von 250 g/kWh dargestellt ist. Die Front wird zum Ende der Vermessung (Iteration 80) sehr genau bestimmt, während zu Beginn die online bestimmten Optima von den tatsächlichen Optima noch weiter entfernt liegen.

In diesem Beispiel reduziert sich die Summe der Abstände zu den 40 Punkten der tatsächlichen Pareto-Fläche von 11.42 in der 27. Iteration auf 2.85 in der letzten Iteration. Die Reduktion wird qualitativ auch im Bild 6.8 a) und b) deutlich. Während in Iteration 27 noch einige Punkte weiter entfernt sind von den Kreuzen, sind die Entfernungen beider in der 80. Iteration wesentlich geringer.

Insgesamt wird deutlich, dass im Rahmen der Vermessung die Pareto-Fläche der Zielfunktion, welche aus der unterschiedlichen Gewichtung der einzelnen Terme resultiert, genau bestimmt wird. Die Modelle sind dadurch für den Einsatz in der anschließenden Anwendung, in welcher je nach Anwendungsfall die benötigten Gewichtungen für ein Optimierungsziel vorgegeben werden, geeignet.

6.4 Vergleich der Online-Vermessung mit einer klassischen Offline-Vermessung

Durch den Einsatz des Online-Verfahrens wird die benötigte Zeit zur Bestimmung von Modellen verkürzt, indem sowohl die benötigte Anzahl an Messpunkten als auch die Dauer der Vermessung reduziert wird.

Um die Dauer der Vermessung zu bewerten, muss berücksichtigt werden, ob die iterative Auswertung länger dauert als die iterative Messung der neuen Messpunkte. Hierfür ist in Bild 6.9 a) der zeitliche Verlauf der Modellbildung und Versuchsplanung (ONLOPOMOT und LOPO-MOT für iterativ gleiche Messdaten) von OTOM dargestellt. Zusätzlich ist in b) die Dauer der eigentlichen Messung in der jeweiligen Iteration dargestellt, wobei beim klassischen Verfahren eine konstante Schaltzeit von 20 s im jeweiligen Messpunkt angenommen wurde.

Dauert die Messung länger als die gesamte Auswertung, entspricht die Dauer der Iteration der Dauer der Messung. Sobald jedoch die Auswertung länger als die Messung dauert, muss der Prüfstand mit der weiteren Messung warten. Die Dauer der Iteration folgt dadurch zur Dauer der Auswertung. Hieraus folgen die Dauer einer Iteration, welche sich aus dem jeweiligen Maximum der Zeiten aus Bild 6.9 a) und b) ergibt, sowie die Gesamtdauer der Vermessung, welche sich aus der Summe der jeweiligen Dauern der einzelnen Iterationen ergibt, zu:

$$T_{\text{Iteration},n} = \max(T_{\text{Modell+Messplan},n}, T_{\text{Messung},n})$$

$$T_{\text{Vermessung}} = \sum_{i=1}^Z T_{\text{Iteration},n} = \sum_{i=1}^Z \max(T_{\text{Modell+Messplan},n}, T_{\text{Messung},n}) \quad (6.3)$$

Dies ist in c) iterativ und in d) kumuliert dargestellt. Werden die Dauern verglichen, zeigt sich der wesentliche Vorteil der iterativen Online-Methode. Wird die Iteration 45 mit 398 Messpunkten betrachtet, benötigt das Online-Verfahren bis dahin insgesamt 75 min, während ein Offline-Verfahren bereits 109 min gedauert hat, was einer Zeitersparnis von ca. 31 % entspricht. Das Offline-Verfahren schafft in den 75 min nur die Vermessung von 298 Messpunkten. Dieser Vorteil der Online-Methode vergrößert sich mit zunehmender Vermessungsdauer. Je länger die Vermessung anhält, desto größer ist der zeitliche Vorteil der Online-Vermessung im Vergleich zum Offline-Verfahren.

Die Dauer der gesamten Auswertung hängt dabei von dem eingesetzten Rechner ab (hier: PC mit 2.40 GHz Prozessor). Durch die Verwendung eines schnelleren Rechners, auf dem OTOM läuft, kann die Auswertung weiter beschleunigt werden. Die Dauer der Messung ist dagegen

unabhängig von der eingesetzten Echtzeithardware. Das Echtzeitsystem muss schnell genug sein, um ORTEM in Echtzeit auszuführen, ansonsten ist eine Online-Vermessung nicht möglich. Ein schnelleres System bringt keine direkten Vorteile.

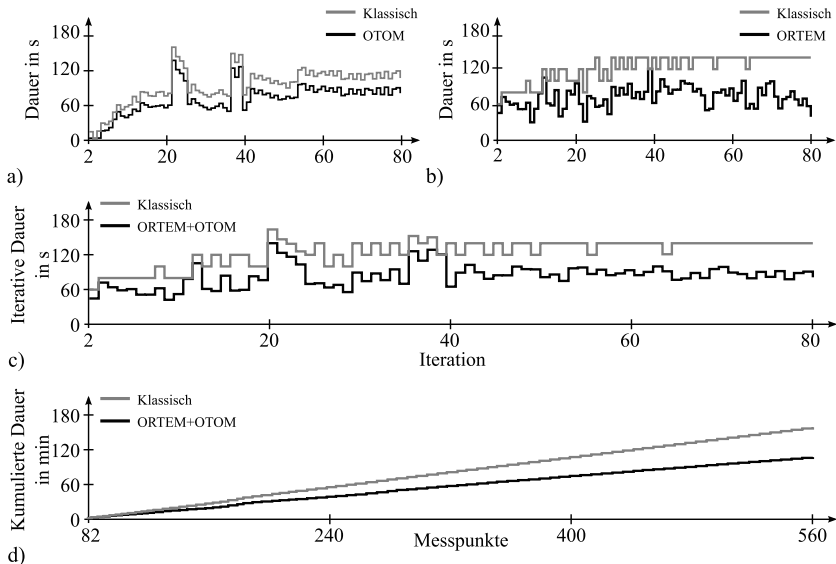


Bild 6.9: Darstellung der zeitlichen Verläufe im Rahmen der Online-Vermessung im Arbeitspunkt ($2250 \text{ min}^{-1} / 18 \frac{\text{mm}^3}{\text{Hub}}$) ohne die erste Iteration (Variationsraumbestimmung). In a) ist die iterative Dauer der Auswertung (Versuchsplanung+Modellbildung) sowie in b) die iterative Dauer der Messung dargestellt. Bei beidem haben die Online-Verfahren einen zeitlichen Vorteil gegenüber klassischen Verfahren. Die resultierende Dauer der Vermessung (jeweiliges Maximum aus Auswertung und Messung) ist in c) iterativ über alle Iterationen und in d) kumuliert über alle Messpunkte dargestellt. Der zeitliche Vorteil der Online-Methode nimmt mit länger dauernder Vermessung zu.

Die Variationsraumbestimmung (ersten 81 Messpunkte) erfolgt gemäß Abschnitt 3.4 in 43 min anstelle 93 min. Die danach ablaufende iterative Vermessung reduziert sich durch den Einsatz einer zielgerichteten Versuchsplanung von 486 auf 317 Messpunkte. Wird die zusätzlich schneller erfolgende Vermessung aufgrund der in Abschnitt 2.3.1 beschriebenen automatischen Schaltzeitbestimmung berücksichtigt, ergibt sich eine um 87 min schnellere Messung (75 min anstelle 162 min). Ein weiterer Vorteil besteht darin, dass im Rahmen der Anwendung der Online-Methodik zur Parametrierung von Sollwertkennfeldern die Optimierung bereits integriert ist. Die Bestimmung und Validierung des Optimums entfällt, wodurch die Dauer (im gezeigten Beispiel ca. 7 min für die Bestimmung und ca. 4 min für die Validierung) weiter verkürzt werden kann.

Die Gesamtersparnis durch den Einsatz der Online-Methode ergibt sich insgesamt zu 148 min bzw. 56 % (118 min anstelle 266 min) für den hier betrachteten Arbeitspunkt. Dies erfolgt auf Kosten einer leicht schlechteren Modellgüte im Vergleich zum Offline-Verfahren, wobei jedoch die vorgegebene Zielgüte erreicht wird. Erfolgt die Vermessung in mehreren Arbeitspunkten (typischerweise 21, siehe z.B.[146]), kann diese durch den Einsatz der Online-Methode um 53 h schneller abgeschlossen werden (41 h anstelle 94 h).

6.5 Zusammenfassung

Die im Rahmen dieser Arbeit entwickelte Online-Methodik zur zielgerichteten und schnellen Online-Vermessung wurde theoretisch entworfen, praktisch umgesetzt und erprobt und steht für den Einsatz im Rahmen von verschiedenen Anwendungsfällen bereit. Das abschließend vorgestellte Beispiel der Vermessung eines lokal statischen Online-Modells mit integrierter globaler Optimierung von 4 Stellgrößen für ein Sollwertkennfeld zeigt das Potential der Online-Vermessung.

Eine zielgerichtete Vermessung ermöglicht das Lösen des Bias-Varianz-Dilemmas, indem die Modelle auf neu hinzukommende Messdaten zielgerichtet adaptiert werden. Hierdurch wird ein überwiegend sinkender Trainings- und Validierungsfehler erreicht.

Im Rahmen eines Online-Verfahrens wird typischerweise (hier im Beispiel zwecks Validierung nicht) die zu erreichende Güte des Modells vorgegeben. Im Vergleich dazu wird beim klassischen Vorgehen die Messpunktanzahl vorgegeben, ohne zu wissen ob diese zu hoch oder zu niedrig für den zugrundeliegenden Anwendungsfall ist. Das Online-Verfahren reduziert durch die auf die Modellgüte zielgerichtete Vermessung die benötigte Anzahl an Messpunkten im Vergleich zum Offline-Verfahren, indem eine zu erreichende Zielgüte vorgegeben wird.

Die in der Vermessung integrierte Optimierung reduziert den Vermessungsaufwand zusätzlich. Die Bestimmung des Optimums erfolgt iterativ, wodurch pro Iteration nur wenige Optimierungsschritte nötig sind. Dies ermöglicht den Einsatz von globalen Optimierern auch bei höherdimensionalen Anwendungsfällen. Weiterhin wird die aus der Optimierung resultierende Pareto-Front automatisch mitvermessen, wodurch eine genaue Validierung dieser möglich ist.

Zusammenfassend erfolgt bei Einsatz des Online-Verfahrens die Vermessung generell schneller als bei Offline-Verfahren. Dem steht der Nachteil einer potentiell leicht schlechteren Modellgüte entgegen, wobei jedoch die zu erreichende Ziel-Modellgüte für das Training und die Validierung vorgegeben werden kann.

7 Zusammenfassung und Ausblick

Die Reduktion von Abgas-Emissionen sowie von Kraftstoffverbrauch gewinnt immer stärkere Bedeutung und führt in der Umsetzung zu einer immer höheren Anzahl an Aktoren und Sensoren und infolgedessen zu einer immensen Zunahme an Variabilitäten. Die Applikation von elektronischen Motorsteuerungen wird folglich zu einer hochdimensionalen Aufgabe und mit der Dimensionalität steigt die Nichtlinearität. Dem entgegen steht der stetig anwachsende Kosten- und Zeitdruck im Rahmen der Entwicklung von neuen Verbrennungsmotoren. Um die Optimierung der Applikation in annehmbarer Zeit und mit akzeptablen Aufwand durchführen zu können, sind sowohl schnelle Vermessungsmethoden als auch modellbasierte Optimierungsmethoden nötig.

Ziel der vorliegenden Arbeit war es, eine zielgerichtete Online-Methodik für die automatische und schnelle Vermessung von Verbrennungsmotormodellen zu entwickeln. Hierdurch kann der Zielkonflikt zwischen Komplexität und Aufwand entschärft werden. Die Methoden wurden implementiert und am Prüfstand erprobt und ermöglichen, den bisherigen Vermessungsprozess von statischen und dynamischen Modellen durch die Verwendung von Online-Methoden zu beschleunigen.

Grundlage einer automatisierten Online-Vermessung stellt die vorgelagerte Bestimmung der zulässigen Aktor-Stellbereiche dar, sodass sowohl ein sicherer Betrieb des Prüfstandes als auch die Vermessung innerhalb vordefinierter Grenzen sichergestellt werden kann. Zu dem Zweck einer beschleunigten Bestimmung der zulässigen Stellbereiche im Vergleich zu bisherigen Methoden wurde eine Prüfstandautomatisierung umgesetzt, welche den Grenzraum automatisch, per geregelter Aktor-Verstellgeschwindigkeit, bestimmt. Die Zeitersparnis durch den Einsatz der geregelten Aktor-Verstellgeschwindigkeit beträgt in etwa $1/3$ der Zeit, welche für die Stellbereichsvermessung mit konstanter Verstellung benötigt wird. Die bei der Stellbereichsvermessung entstehenden Hüllpunkte wurden im Anschluss mithilfe der neu entwickelten Methodik zur einfachen Beschreibung von konkaven Hüllen genutzt, um nicht zulässige Stellbereiche bei der automatischen Vermessung auszuschließen. Durch die einfache Beschreibung der konkaven Hüllen kann die Bestimmung vollständig in die Vermessung integriert werden, wodurch eine nachträgliche Auswertung am Arbeitsplatz gänzlich entfällt. Weitere Prüfstandszeit kann eingespart werden, indem Variationsräume von Zwischenarbeitspunkten aus den Räumen der umliegenden Arbeitspunkte interpoliert werden, anstelle diese aufwändig neu zu vermessen.

Im Anschluss an die Bestimmung der zulässigen Stellbereiche wurden zur geeigneten Vermessung des zugrundeliegenden Prozesses aktive lernende Versuchspläne entwickelt. Diese bewerten sowohl die Raumfüllung als auch die aktuell bekannten Modelleigenschaften und platzieren neue Messpunkte an Stellen, an denen der Informationsgewinn für die folgende Modellbildung und/oder Stellgrößentheorie am größten ist. Infolgedessen können auch Prozesse, welche zu Beginn völlig unbekannt sind, zielgerichtet vermessen werden. Ergänzt wird die Versuchsplanung der dynamischen Vermessung um eine automatische signalbasierte Bestimmung der Abtastzeit, damit das Maximum an Information aus den Signalen extrahiert werden kann. Durch

den Einsatz der vorgestellten Methoden zur Maximierung des Informationsgehaltes bei der Versuchsplanung kann eine schnellere Konvergenz gegen die maximal erreichbare Modellgüte erreicht werden. Dadurch wird bereits mit weniger Messpunkten ein besseres Modell bestimmt. Die Zeitersparnis ist hierbei stark von dem modellierten Prozess sowie der gewünschten Modellgüte abhängig und liegt bei dem vorgestellten Beispiel bei ca. 50% weniger Messpunkten. Mithilfe eines prozessangepassten deterministischen Rechtecksignals, welches tiefe Frequenzen stark anregt, kann der Prozess flexibel angeregt werden. In Folge kann mit einer ähnlichen Signallänge wie z.B. die eines APRBS, eine höhere statische sowie dynamische Güte des identifizierten Modells erreicht werden.

Nach Aufzeichnung der Messdaten erfolgt die Modellbildung, welche für den Einsatz im Rahmen der Online-Vermessung zu der online-fähigen ONLOPOMOT-Modellbildung erweitert wurde. ONLOPOMOT ist ein adaptives lokal polynomiales Netzmodell, welches eine iterative Adaption der lokalen Modellparameter, der Modellstruktur sowie der Netzmodellteilung beinhaltet. Die Regressorauswahl erfolgt kontinuierlich durch die Kombination der LASSO-Parameterschätzung mit einer Erhöhung der maximal zulässigen Regressorordnung bei Bedarf. Die Bestimmung der passenden Regressoren wird durch die Teilungsanpassung des Netzmodells ergänzt, wobei neben der klassischen Vergrößerung bei Bedarf auch eine Verkleinerung des Netzmodells enthalten sein kann. Wesentliche Eigenschaft der Online-Modellbildung ist, dass das Bias-Varianz-Dilemma, aufgrund der zunehmenden Anzahl an Messpunkten bei gleichzeitig steigender Modellkomplexität, entschärft wird. Hierdurch kann sowohl die Trainings- als auch die Validierungsgüte mit zunehmender Messdauer verbessert werden. Zusätzlich kann durch die bedarfsgerechte Adaption des Modells der zeitliche Gesamtaufwand zur Modellbildung reduziert werden. Während dieser bei der Offline-Modellbildung exponentiell mit der Anzahl an Datenpunkten zunimmt, wird dessen Zunahme bei der Online-Modellbildung mit weiteren Messpunkten immer kleiner. Zusätzlich wurde eine Methode zur Erweiterung der klassischen Modellbildung um die Modellierung der Richtungsabhängigkeit von Prozessen dargestellt. Durch die Ergänzung der Regressormatrix zur Parameterschätzung um eine richtungsabhängige Schaltfunktion wird die automatische Identifikation der Richtung von statischen oder dynamischen Prozessen ermöglicht. Hierdurch wird die Güte von Modellen, welche einen Prozess mit richtungsabhängigem Verhalten (z.B. pneumatischer VTG-Aktor) nachbilden sollen, signifikant verbessert werden.

Abschließend wurden die entwickelten Methoden in eine durchgehende Gesamtstrategie eingebettet. Die sich daraus ergebenden neuen Anforderungen an den zugrundeliegenden Prüfstand sowie dessen Automatisierung wurden erläutert. Alle im Rahmen dieser Arbeit entwickelten Algorithmen wurden in die Toolboxes OTOM und ORTEM integriert. Erstere entspricht der Weiterentwicklung von Offline-DoE-Tools zum Einsatz in der Online-Methodik während letztere eine neu entwickelte Prüfstandsautomatisierung darstellen. Durch die Kombination beider wurde eine zielgerichtete automatisierte Online-Vermessung an einem Prüfstand mit Echtzeithardware realisiert. Hierdurch wird der gesamte Aufwand zur Nachbereitung, welcher bei klassischen Methoden im Anschluss anfällt, eingespart.

Die in dieser Arbeit entwickelte Gesamtmethodik wurde nicht nur aus einem theoretischen Gesichtspunkt untersucht, sondern wurde auch für den praktischen Einsatz umgesetzt und steht zur Weiterverwendung bereit. Es wurde gezeigt, dass durch die Kombination der Variationsraumbestimmung, der Versuchsplanung und der Identifikation in eine Gesamtstrategie zur Online-Vermessung der Vermessungsprozess vollständig automatisiert werden kann. Die Bestimmung der Modelle erfolgt dadurch unabhängig von dem Wissen des Applikateurs/Prüfstandsfahrers. Durch die zielgerichtete und schnellen Identifikation können Verbrennungsmotormodelle effizient vermessen werden. Die resultierende Reduktion des Gesamtaufwands ist stark vom zugrundeliegenden Prozess bzw. der Anwendung abhängig. In jedem Fall wird der Aufwand für die Auswertung und Modellbildung eingespart. Darüber hinaus können in verschiedenen Anwendungsfällen jeweils ca. 1/3 der Messzeit sowie der Messpunkte zur Erreichung einer gewünschten Modellgüte eingespart werden.

Ausblick

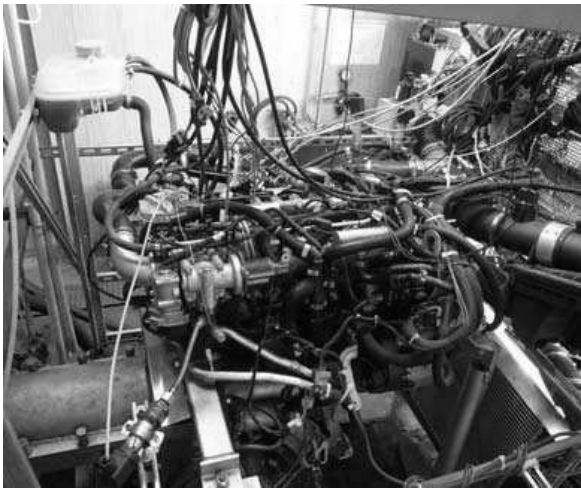
Die Anwendbarkeit der umgesetzten Online-Methodik wurde exemplarisch für verschiedene Beispiele untersucht und zeigt ein großes Potential für den Einsatz bei modellbasierten Methoden. Die vorgestellte Vorgehensweise kann für den Einsatz bei weiteren Anwendungsfällen ergänzt werden, wie z.B. für den modellbasierten Entwurf von Steuerungen und Regelungen oder für modellbasierte Diagnosen. Hierfür muss die Online-Methodik um die spezifischen Randbedingungen, die jede Anwendung mit sich bringt, ergänzt werden. Durch einen somit ermöglichten automatisierten Entwurf kann der Entwicklungsprozess noch weiter verkürzt werden.

Der zeitaufwändigste Aspekt in der aktuellen Methodik ist die Optimierung. Hier kommen bisher klassische Optimierungsverfahren zum Einsatz, wobei deren Startpunkte iterativ verbessert werden. Durch die Weiterentwicklung der Optimierung hin zu einer Online-Optimierung, welche sich parallel zu der iterativen Vermessung automatisiert adaptiert, kann eine weitere Beschleunigung der vorgestellten Methodik erreicht werden.

Weiterhin ermöglicht der iterative Vermessungsablauf sowie die modulare Struktur der Online-Auswertung prinzipiell den Einsatz von Offline-Methoden (z.B. Support-Vektor-Modelle, etc.), welche ohne signifikante Änderung in die Methodik integriert werden können. Nachteilig ist hierbei immer, dass diese Methoden selten auf Effizienz ausgelegt sind. In der Entwicklung dieser Offline-Methoden hin zum Online-Einsatz steckt weiteres Potential, neue Anwendungsbereiche zu erschließen, bei denen bislang diese Offline-Methoden zum Einsatz kommen.

A Motorenprüfstand und Versuchsträger

Der für die Verifizierung der Methodik zum Einsatz gekommene Verbrennungsmotor ist ein Opel/Fiat 1.9l Common-Rail Dieselmotor. Dieser verfügt über ein Turbolader mit variabler Turbinengeometrie sowie ein Hochdruck- und ein Niederdruck-AGR-Ventil. In Bild A.1 ist ein Bild des Motors mit der Fülle an zusätzlicher Sensorik sowie die wesentlichen Kenndaten dargestellt. Neben den Seriensensoren sind an diesem Motor zusätzlich Sensoren für verschiedene Temperaturen, Drücke, Emissionen (Ruß, Opazität und NO_x) sowie das Moment verbaut. Ein Überblick aller möglichen Stell- und Messgrößen ist in Bild A.2 veranschaulicht.



Hersteller	GM/Opel & Fiat	Bohrung	82 mm
Modell	Z19DTH	Kurbelwellenradius	45,2 mm
Steuergerät	Bosch EDC16 C9	Pleuellänge	145 mm
Typ	4-Takt Diesel	Polytropenexponent	1.37
Anzahl Zylinder	4	Max. Leistung	110 kW (4000 min ⁻¹)
Anzahl Ventile	16	Max. Moment	240 Nm (2300 min ⁻¹)
Hubraum	1909.6 cm ³	Max. Drehzahl	5100 min ⁻¹
Verdichtung	17.5		

Bild A.1: Bild des Versuchsmotors mitsamt Asynchronmaschine als Bremse sowie Sensorik zur Messung der verschiedenen Größen sowie Zusammenstellung der wichtigsten technischen Daten des Versuchsträgers [146].

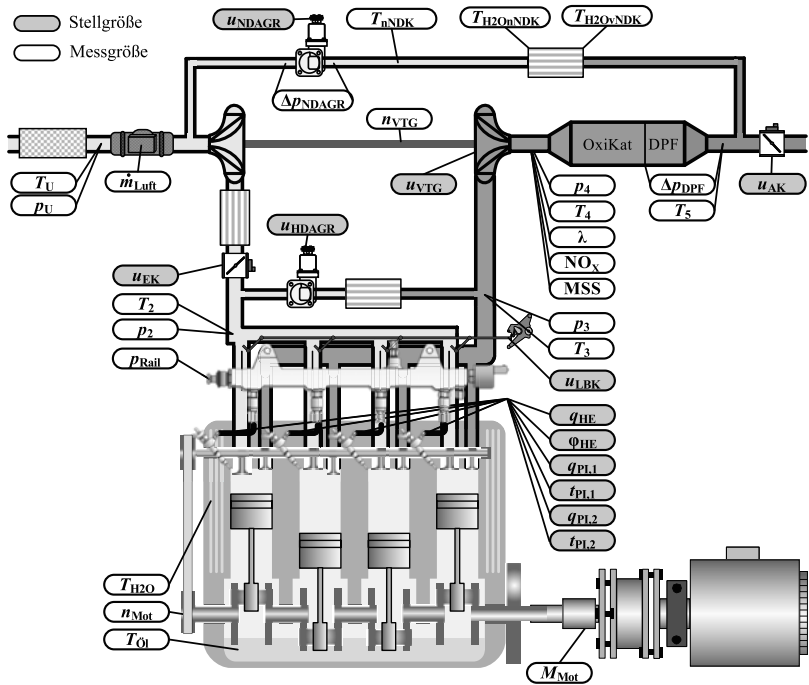


Bild A.2: Überblick aller Stell- und Messgrößen des Versuchsmotors.

B Software Dokumentation

Alle im Rahmen dieser Arbeit entwickelten Algorithmen wurden in Matlab 2011b umgesetzt und in das Software-Paket *Online-Methoden Toolbox* integriert. Dabei wurden die Algorithmen nicht nur aus theoretischen Gesichtspunkten untersucht, sondern auch im praktischen Einsatz (unter Matlab 2011b und dSPACE 7.2) validiert und stehen zur Weiterverwendung bereit. Zusätzlich ist ein Testmodus integriert, welcher eine iterative Vermessung simuliert.

Als Systemumgebung wird ein PC vorausgesetzt, welcher mit einer Matlab-Lizenz ausgestattet ist und über die Kommunikation mit einer Echtzeithardware verfügt, welche einen Zugriff auf alle benötigten Aktoren und Sensoren hat. Das Software-Paket besteht aus den beiden Toolboxes *OTOM* (online target oriented measurement) und *ORTEM* (online real-time engine measurement).

ORTEM

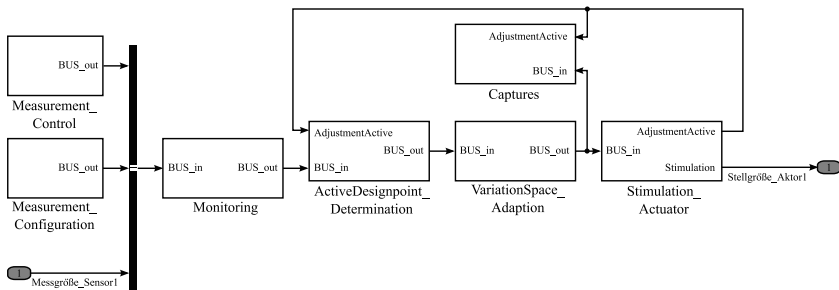


Bild B.1: Höchste Hierarchieebene des ORTEM-Simulinkmodells.

Das Online-Vermessungssystem ORTEM ist ein Simulink-Modell (insgesamt 51 Module), welches mithilfe eines entsprechenden Kompilierers für den Einsatz auf Echtzeithardware gewandelt werden kann. Es ermöglicht eine echtzeitfähige Verstellung der Aktoren bei gleichzeitiger Limit-Überwachung. Durch die Erweiterung um die automatische Adaption der Aktor-Verstellgeschwindigkeit sowie den automatischen Übergang zum nächsten Messplanpunkt bei Stationärerwerterkennung wird insgesamt die Messzeit im Vergleich zu klassischen Prüfstandautomatisierungstools reduziert.

In Bild B.1 ist ORTEM auf der höchsten Hierarchieebene dargestellt. Die sich aus den 7 Modulen ergebende Modellhierarchie folgt zu:

1. Measurement_Control
 - (a) Activation_Actuators

- (b) Activation_Adaption
- (c) Activation_RangeDetection
- (d) Activation_StationaryDetection
- 2. Measurement_Configuration
 - (a) Design_Configuration
 - (b) Sensors_Configuration
 - (c) VariationSpace_AdaptionConfiguration
- 3. Monitoring
- 4. ActiveDesignpoint_Determination
 - (a) Counter_Determination
 - (b) Limitviolation_hard
 - (c) Switchtime_Determination
- 5. VariationSpace_Adaption
 - (a) AdaptionFactor_Determination
 - (b) Adaption_BorderRegion
- 6. Stimulation_Actuator
- 7. Captures
 - (a) DataCapture_Modelling
 - (b) DataCapture_VariationSpace
 - (c) Trigger_Captures

In den Blöcken *Measurement_Control* und *Measurement_Configuration* erfolgt die Steuerung des Echtzeitsystems und Parametrierung der Vermessung in jeder Iteration, indem die entsprechenden Parameter durch OTOM geschrieben werden. Das *Monitoring* enthält die Überwachung aller Sensoren im Rahmen der Variationsraumvermessung. Im Block *ActiveDesignpoint_Determination* ist der Teil des Echtzeit-Modells umgesetzt, welcher die echtzeitfähige Aktorverstellung realisiert. Die Regelung der Verstellgeschwindigkeit an die Variationsraumgrenzen erfolgt in *VariationSpace_Adaption*. Die beiden Blöcke *Stimulation_Actuator* und *Captures* stellen die Schnittstellen von ORTEM dar, sowohl zu den Aktoren des Verbrennungsmotors als auch zur Datenaufzeichnung auf der Festplatte des Computers.

OTOM

Das Online-Auswertungssystem OTOM ist in m-Code (insgesamt 112 m-Files) umgesetzt und nutzt Messdaten sowie identifizierte Modelle zur aktiv lernenden Adaption von Messplänen und Modellen. Dabei werden diese solange iterativ angepasst und neue Messdaten vermessen, bis die Zielgüte erreicht ist. Infolgedessen kann sowohl eine maximale Verbesserung des Modells im Hinblick auf Bias und Varianz erreicht werden, als auch eine unnötige Vermessung von Messpunkten ohne wesentliche Prozessinformation vermieden werden, was wesentliche Messzeit einspart. Für eine detaillierte Beschreibung der einzelnen Funktionen bzw. m-Files sei an dieser Stelle auf den kommentierten Quellcode [93] verwiesen.

Anleitung zum Einsatz

Zur Installation der Toolbox und zum Starten einer Vermessung müssen

- in der m-Datei *OnlineMeasurement.m* der OTOM-Toolbox die Variable *Path* an die jeweilige Rechnerumgebung angepasst werden,
- in der m-Datei *configTestbed.m* verschiedene Prüfstandsdaten (Messdatenerfassungsfrequenz, Pfade zu Toolboxen, zu speichernde Dateinamen, etc.) parametrisiert werden,
- in der m-Datei *configMeasurement.m* verschiedene Vermessungsdaten (Variationsraumvermessung, Versuchsplanung, Modellbildung, etc.) parametrisiert werden,
- in der m-Datei *configMlib.m* verschiedene Schnittstellendaten (Sensorpfade der Echtzeithardware, Aktorpfade der Echtzeithardware, etc.) parametrisiert werden,
- ORTEM um die Sensoren und Aktoren des Prüfstands ergänzt und kompiliert werden,
- das kompilierte Modell auf die vorliegende Echtzeithardware geladen und gestartet werden.

Zum Starten der iterativen Online-Vermessung muss abschließend die m-Datei *OnlineMeasurement.m* ausgeführt werden. Sobald dies erfolgt, läuft die Vermessung selbstständig bis zum Erreichen des Abbruchkriteriums.

Um den Testmodus zu aktivieren, muss im zweiten Schritt bei der Parametrierung der Prüfstandsdaten die Variable *SettingsTestbed.Type* auf Offline gestellt werden. Dieser Modus benötigt keine Echtzeithardware, sondern simuliert die iterative Vermessung mithilfe eines hinterlegten Beispielprozesses.

C Höherdimensionale konkave Variationsräume

Das Vorgehen zur Bestimmung von hochdimensionalen konkaven Variationsräumen entspricht dem aus Abschnitt 3.5, jedoch mit auf höhere Dimensionen erweiterten Berechnungsvorschriften, welche im folgenden Beschrieben werden.

Bestimmung hochdimensionaler Variationsräume

Zur Bestimmung der fälschlicherweise bei der Hüllbildung ignorierten, innen-liegenden O Stellgrößenkombinationen $\mathbf{u}_i = (u_{i,1}, u_{i,2}, \dots, u_{i,P})$ mit $i = 1, \dots, O$ und der Dimension P müssen diese in Kugelkoordinaten transformiert werden. Die in Formel 3.18 definierte Transformation erweitert sich im bei P -dimensionalen Räumen zu:

$$\begin{bmatrix} u_{1,1} & \dots & u_{P,1} \\ \vdots & \ddots & \vdots \\ u_{1,O} & \dots & u_{P,O} \end{bmatrix} \Rightarrow \begin{bmatrix} \varphi_{1,1} & \dots & \varphi_{P-1,1} \\ \vdots & \ddots & \vdots \\ \varphi_{1,O} & \dots & \varphi_{P-1,O} \end{bmatrix} \star \begin{bmatrix} r_1 \\ \vdots \\ r_O \end{bmatrix} \quad (\text{C.1})$$

Durch die Transformation wird jede P -dimensionale Stellgrößenkombination \mathbf{u}_i durch einen Radius r_i und $p = 1, \dots, P - 1$ Winkel $\varphi_{p,i}$ beschrieben. Somit entsprechen im hochdimensionalen Fall gleiche Richtungen gleichen Winkeln φ_i und das Vorgehen zur Bestimmung der radial verschobenen Stellgrößenkombination und fälschlicherweise hinzugenommen Simplexe entspricht dem Vorgehen im zweidimensionalen Fall.

Eine Triangulation hat die Eigenschaft, dass der Raum in mehrere P -Dimensionale Simplexe zerlegt wird, wobei jedes davon durch $P + 1$ -Kanten definiert ist. Bei der Bestimmung der konkaven Hülle wird diese Eigenschaft verwendet. Bei N Simplexen ergibt sich somit ein Datensatz von $N \cdot (P + 1)$ Kanten. Dieser beinhaltet dabei sowohl mehrmals vorkommende Kanten (jeweilige Kanten von zwei aneinanderliegenden Simplexen) als auch einmalig vorkommende Kanten (Kante eines Simplex am Rand der Triangulation). Alle im Datensatz nur einmalig vorkommenden Kanten definieren somit die äußere konkave Hülle.

Mithilfe der in Abschnitt 3.5 beschriebenen und hier für höherer Dimensionen erweiterten Vorgehensweise können hochdimensionale konkave Räume dargestellt und in der integrierten Gesamtmethodik berücksichtigt werden, ohne dass ein manuelles Anpassen notwendig ist.

Interpolation hochdimensionaler Variationsräume

Auch das Vorgehen zur Interpolation hochdimensionaler Variationsräume ($P \geq 5$) entspricht dem aus Abschnitt 3.5, jedoch mit erweiterter Berechnungsvorschrift. Für jede maximal zulässige Stellgrößenkombinationen \mathbf{u}_i (Stützstelle) muss in allen $j = 1 \dots J$ umliegenden Hüllen der sich ergebende Radius $r_{VR,j,i}$ bei gleicher Stützstellenrichtung φ_j bestimmt werden.

Im ersten Schritt werden dabei die kartesischen Koordinaten der Stützstellen in ihre Kugelkoordinaten transformiert. Im Fall von zwei globalen Stellgrößen und $P - 2$ lokalen Stellgrößen erfolgt die Interpolation bilinear. Dadurch müssen bei den 4 (anstelle 2 im 3-dimensionalen Fall) benachbarten Variationsräumen VR_1, \dots, VR_4 gegenseitig fehlenden Stützstellen ergänzt werden. Die hierfür benötigte Berechnung des Schnittpunktes erfolgt mithilfe einer Geraden- und einer $(P - 2)$ -dimensionalen Ebenengleichungen.

Für jeden beliebigen Punkt $s_{VRj,i}$ auf einer Strecke zwischen zwei Punkten z_{VRj} und $m_{VRj,i}$ gilt folgende Gleichung in Zweipunkteform:

$$s_{VRj,i} = z_{VRj} + a \cdot (m_{VRj,i} - z_{VRj}). \quad (C.2)$$

Dabei sind im Versuchsraum z_{VRj} der Zentralpunkt und $m_{VRj,i}$ der Maximalpunkt, welcher bei Darstellung in Kugelkoordinaten dem durch Verlängerung des Radius r_{VRj} bis zur Grenze des Stellbereichs entstehenden Punktes entspricht.

Zusätzlich kann für jede durch die $(P - 2)$ Stützstellen $u_{VRj,i}, \dots, u_{VRj,i+P-2}$ (jede Stützstelle ist dabei $(P - 2)$ -dimensional) aufgespannte Facette (Seitenfläche der $(P - 2)$ -dimensionalen Hülle), die Facettengleichung in Parameterform aufgestellt werden:

$$s_{VRj,i} = u_{VRj,i} + b_1 \cdot (u_{VRj,i+1} - u_{VRj,i}) + b_2 \cdot (u_{VRj,i+2} - u_{VRj,i}) + \dots + b_{P-3} \cdot (u_{VRj,i+P-3} - u_{VRj,i}) \quad (C.3)$$

Durch Gleichsetzen und lösen der beiden Gleichungen mithilfe der Matrixinversion

$$\begin{bmatrix} a \\ b_1 \\ \vdots \\ b_{P-3} \end{bmatrix} = \begin{pmatrix} \begin{bmatrix} z_{VRj} - m_{VRj,i} \\ u_{VRj,i+1} - u_{VRj,i} \\ u_{VRj,i+2} - u_{VRj,i} \\ \vdots \\ u_{VRj,i+P-3} - u_{VRj,i} \end{bmatrix}^T \end{pmatrix}^{-1} \cdot [z_{VRj} - u_{VRj,i}] \quad (C.4)$$

kann der Schnittpunkt der Strecke $\overline{z_{VRj}m_{VRj,i}}$ mit jeder durch $\overline{u_{VRj,i}u_{VRj,i+1} \dots u_{VRj,i+P-3}}$ aufgespannten Facette bestimmt werden. Dabei ist zu beachten, dass nur die Lösung, für die

$$a, b_1, \dots, b_{P-3} \in [0, 1] \text{ und } 0 \leq \sum (b_1 + \dots + b_{P-3}) \leq 1 \quad (C.5)$$

gilt, den gesuchten Schnittpunkt $s_{VRj,i}$ liefert und somit die fehlende Stützstelle in dem Variationsraum definiert.

Nachdem die umliegenden Räume mithilfe des beschriebenen Vorgehens um die fehlenden Stützstellen der gleichen Richtung ergänzt worden sind, kann der Radius der umliegenden Stützstellen gleicher Richtung bilinear interpoliert werden. Im Vergleich zu kartesischen Koordinaten, bei denen alle Dimensionen einzeln interpoliert werden müssten, reduziert sich der Rechenaufwand um das $(P - 1)$ -fache. Der aus den vier umliegenden Variationsräumen VR_1, \dots, VR_4

interpolierte Radius $r_{VR5,i}$ des Variationsraums VR_5 berechnet sich gemäß der bilinearen Interpolation der globalen Stellgrößen $u_{P-1,VRj}$ und $u_{P,VRj}$ zu:

$$\begin{aligned}
 r_{VR5,i}(u_{P-1,VR5}, u_{P,VR5}) = & r_{VR1,i} \frac{(u_{P-1,VR5} - u_{P-1,VR1})(u_{P,VR5} - u_{P,VR1})}{(u_{P-1,VR2} - u_{P-1,VR1})(u_{P,VR4} - u_{P,VR1})} + \\
 & r_{VR2,i} \frac{(u_{P-1,VR2} - u_{P-1,VR5})(u_{P,VR5} - u_{P,VR1})}{(u_{P-1,VR2} - u_{P-1,VR1})(u_{P,VR4} - u_{P,VR1})} + \\
 & r_{VR3,i} \frac{(u_{P-1,VR3} - u_{P-1,VR5})(u_{P,VR3} - u_{P,VR5})}{(u_{P-1,VR2} - u_{P-1,VR1})(u_{P,VR4} - u_{P,VR1})} + \\
 & r_{VR4,i} \frac{(u_{P-1,VR5} - u_{P-1,VR4})(u_{P,VR4} - u_{P,VR5})}{(u_{P-1,VR2} - u_{P-1,VR1})(u_{P,VR4} - u_{P,VR1})}.
 \end{aligned} \tag{C.6}$$

Entsprechend können alle interpolierten Radien $r_{VR5,i}$ für jede gesuchte Stützstelle i bestimmt werden. Zusammen mit den zugehörigen Winkeln definiert der interpolierte Radius den maximal möglichen Stellbereich bzw. Variationsraum in dem gesuchten Zwischenraum VR_5 .

D Dynamische Variationsräume

Neben der Bestimmung und Darstellung von statischen Variationsräumen kann auch der dynamische Variationsraum ermittelt werden. Ausgehend von dem statischen Variationsraum kann dabei angenommen werden, dass die berücksichtigten Ausgangsgrößen bei einer dynamischen Verstellung der Aktoren auf die Stellgrößenkombinationen der Hüllpunkte nach Abklingen aller dynamischen Einschwingvorgänge ihren statischen Grenzwert erreichen.

Ein dynamischer Variationsraum wird definiert durch Gütekriterien, welche aus der Optimierung von Regelkreisen [65] stammen. Im Gegensatz zu einer Regelung ist für den dynamischen Variationsraum nur die maximale Überschwingweite von Interesse. Weitere Kenngrößen wie z.B. die summierte Regelabweichung oder die Anregelzeit sind nicht relevant. Um die Kenngröße eines Hüllpunktes zu ermitteln, wird eine dynamische Anregung des entsprechenden Hüllrasterpunktes mit einem Anregungssignal aus Abschnitt 4.1.4 durchgeführt.

Bild D.1 zeigt die Signalverläufe zur Bestimmung des dynamischen Variationsraums für das folgende Hammerstein-Modell mit zwei Eingangsgrößenverläufen und einer Ausgangsgröße sowie $T_0 = 0.01$ s:

$$\begin{aligned}
 \tilde{u}(t) &= 2 + 2u_1(t) + 4u_1(t)^2 + 2u_2(t)^2 + 3u_2(t)^3 \\
 y(t) &= \frac{1}{0.25s + 0.2s + 1} \tilde{u}(t) \\
 &\Downarrow \\
 y(z) &= \frac{0.0001995z + 0.0001989}{z^2 - 1.992z + 0.992} \tilde{u}(z)
 \end{aligned} \tag{D.1}$$

Der Grenzraum wird anhand der Limitierung der Ausgangsgröße auf Werte zwischen $3 \leq y \leq 7$ festgelegt. In Bild D.1 a) sowie b) sind die Signalverläufe der Ein- und Ausgangsgrößen für eine unterschiedliche Messpunktfolgenfolge bei sprungförmiger Anregung dargestellt. Das dynamische Überspringen und somit der dynamische Variationsraum bei einer Punktsortierung nach minimalem Abstand (a) ist kleiner als bei einer Messpunktsortierung mit maximalem Abstand. Zusätzlich sind in c) und d) die jeweiligen statischen Variationsräume mit der zugehörigen limitierten Ausgangsgröße y dargestellt, wobei die Ausgangsgrößen für den statischen und dynamischen Fall abgebildet sind. Die unterschiedliche Überschreitung der Grenzwerte wird deutlich, wobei die Grenzwerte des statischen Variationsraums nur für einen kurzen Zeitraum überschritten werden.

Die Vermessung des dynamischen Variationsraums würde eine Reduktion der maximal zulässigen Stellgrößen soweit erforderlich machen, bis die maximale Überschwingweite innerhalb der statischen Grenzen liegt. Dabei ist zu beachten, dass der dynamische Variationsraum nur für weiche Grenzen ermittelt werden kann, da ein Überspringen einer harten Grenze eine Beschädigung des Motors zur Folge haben könnte. Theoretisch wäre es möglich, einen dynamischen Variationsraum, aus dem Informationen für die Vermessung (z.B. Anpassung des dynamischen Anregungssignals) gefolgert werden können, so zu bestimmen. Um verschiedene

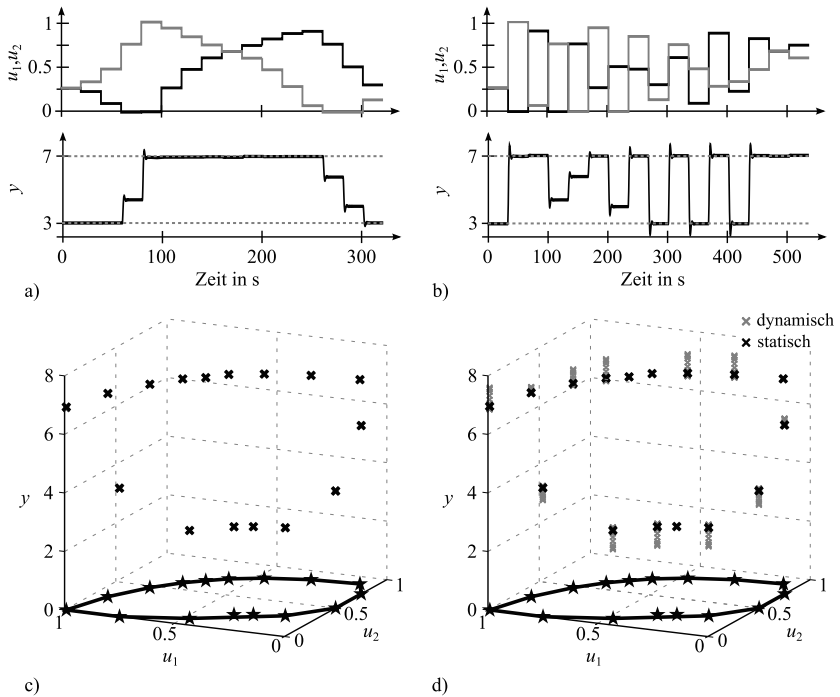


Bild D.1: Signalverläufe der relevanten Stell- und Ausgangsgrößen zur Bestimmung des a) statischen sowie b) dynamischen Variationsraums. Die resultierenden zulässigen Stellbereiche sind in c) für den statischen bzw. in d) für den dynamischen Variationsraum dargestellt

Dynamiken in verschiedenen Stellbereichen erfassen zu können, müsste dafür die dynamische Reaktion aller Hüllpunkte bei Anregung mit verschiedenen Signalen aus verschiedenen Richtungen mit verschiedenen Amplituden erfasst werden. Der hierfür benötigte Aufwand würde exponentiell steigen und wäre schon bei zwei Stellgrößen nicht mehr sinnvoll durchführbar.

E Bestimmung der Abtastzeit ohne Prozesswissen bei dynamischen Messungen

Die Faustformeln aus Abschnitt 4.1.1 haben den Nachteil, auf Tiefpassprozesse mit bekannten Zeitkonstanten beschränkt zu sein. Im Rahmen der Online-Schätzung ist jedoch der zugrundeliegende Prozess sowie dessen Zeitkonstante unbekannt, wodurch die passende Parametrierung der Abtastzeit vor Beginn der Vermessung nicht möglich ist. Aus diesem Grund wird im Rahmen der Online-Methodik ein heuristischer Ansatz umgesetzt.

Die Abtastzeitkonstante richtet sich nach dem Prozessverhalten und sollte so klein wie möglich gewählt werden, unter Berücksichtigung eines ausreichend großen Signal-Rausch-Verhältnisses (SNR, signal to noise ratio) sowie der auftretenden numerischen Probleme bei viel zu kleinen Abtastzeiten [70]. Daraus abgeleitet wird für den heuristischen Ansatz eine Einstellregel angewandt, welche das Signal-Rausch-Verhältnis in seiner alternativen Definition aus der Bildverarbeitung nicht als Verhältnis der Signalleistungen, sondern als Verhältnis der maximal auftretenden Signaländerung Δy und der Standardabweichung $\sigma_{y,\text{init}}$ betrachtet:

$$\begin{aligned} \text{SNR}(T_0) &= \frac{\max_{k=1,\dots,N-1} \left(\Delta y(kT_0) \right)}{\sigma_{y,\text{init}}} \\ &= \frac{\max_{k=1,\dots,N-1} \left(y(kT_0 + T_0) - y(kT_0) \right)}{\sigma_{y,\text{init}}} \approx 6 \dots 8 \end{aligned} \quad (\text{E.1})$$

Der Zähler stellt die maximal auftretende Differenz der gemessenen Ausgangsgröße zwischen zwei aufeinanderfolgenden Samples im betrachteten Datensatz dar und ist abhängig von der Abtastzeit. Unter der Annahme einer konstanten Störung der gemessenen Ausgangsgröße kann deren Standardabweichung $\sigma_{y,\text{Sensor}}$ in einem konstanten Arbeitspunkt bestimmt werden, wobei diese annähernd unabhängig von der Abtastzeit T_0 ist.

Eine größere Abtastzeit führt zu einem größeren $\Delta y(kT_0)$ und umgekehrt. Eine kleine Abtastzeit führt dazu, dass die Differenz zwischen zwei aufeinanderfolgenden Samples klein ist, im Extremfall kleiner als die Standardabweichung des Rauschens. Eine Abtastung mit einer solchen Abtastzeit würde somit vor allem das Rauschen darstellen. Eine größer werdende Abtastzeit führt dazu, dass ab einer bestimmten Schwelle der Zähler größer als die Standardabweichung des Rauschens wird und somit nicht mehr das Rauschen, sondern vor allem die Prozesseffekte dargestellt werden. Wird die Abtastzeit jedoch weiter vergrößert, erhöht sich zwar der Zähler und folglich das SNR, allerdings werden auch die dynamischen Effekte des Prozesses schlechter abgebildet. Eine geeignete Abtastzeit, welche groß genug ist, um nicht mehr allein das Rauschen darzustellen, jedoch klein genug ist, um die Dynamik zu zeigen, ergibt sich für ein Signal-Rausch-Verhältnis im Bereich $6 \dots 8$.

In Bild E.1 sind für einen Beispielprozess mit mehreren Zeitkonstanten

$$G(s) = \frac{-8s + 1}{10s^2 + 7s + 1} \quad (\text{E.2})$$

mit einem überlagerten Rauschen mit der Standardabweichung von $\sigma = 0.05$ der diskrete Signalverlauf der Sprungantwort für verschiedene Abtastzeiten T_0 dargestellt. Für die drei Abtastzeiten $T_0 = 0.1 \text{ s}$, $T_0 = 1.5 \text{ s}$ und $T_0 = 4.5 \text{ s}$ ergeben sich die folgenden Signal-Rausch-Verhältnisse:

$$\begin{aligned} \text{SNR}(T_0 = 0.1 \text{ s}) &= 4.6 \\ \text{SNR}(T_0 = 1.5 \text{ s}) &= 6.765 \\ \text{SNR}(T_0 = 4.5 \text{ s}) &= 12.12 \end{aligned} \quad (\text{E.3})$$

Aus dem Graphen in E.1 wird deutlich, dass mit der Wahl von $T_0 = 1.5 \text{ s}$ der Prozess geeignet abgetastet wird. Die kleinere Abtastzeit führt zu einer starken Abtastung des Prozessrauschens, wodurch sich im Verlauf der Identifikation numerische Probleme ergeben würden [70]. Wird die Abtastzeit zu groß gewählt, können die dynamischen Effekte mit kleiner Zeitkonstante nicht ausreichend dargestellt werden, wodurch das identifizierte Modell die relevante Prozessdynamik nicht vollständig beinhalten würde.

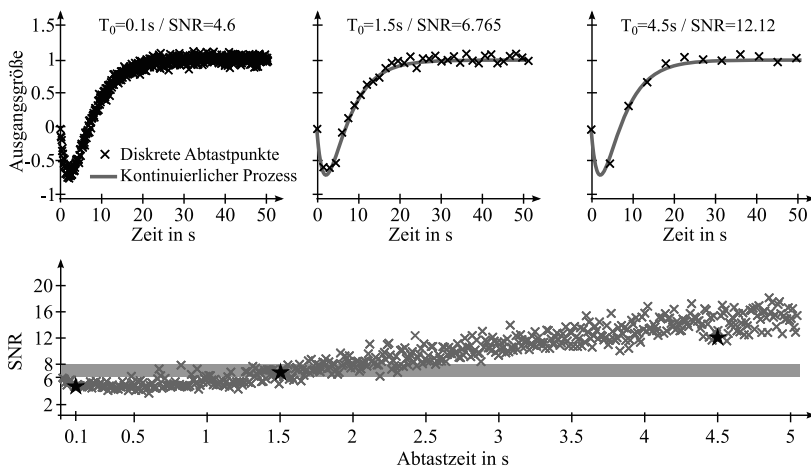


Bild E.1: Darstellung eines PDT₂-Prozesses mit überlagertem Rauschen für verschiedene Abtastzeiten. Eine zu kleine Abtastzeit führt zwar zu einem Informationsgewinn über das Rauschen, jedoch zu numerischen Problemen im Rahmen der Identifikation. Eine zu große Abtastzeit führt dagegen zu dem Verlust der Information über die Prozessdynamik.

Als weiteres Beispiel sind die kontinuierlichen und diskreten Sprungantworten der Beispielprozesse aus [173] in Bild E.2 dargestellt, jeweils überlagert mit demselben Rauschen von $\sigma = 0.05$. Die Abtastzeit wurde für $\text{SNR} > 7$ zu $T_0 = 0.67 \text{ s}$ bestimmt. Es wird deutlich, dass

mit der SNR-basierten Einstellung der Abtastzeit alle dynamischen Effekte des Prozesses ausreichend abdeckt werden, ohne dass Vorwissen über den zugrundeliegenden Prozess vorhanden sein muss, was bei den Einstellregeln basierend auf T_{95} bzw. T_{\min} der Fall ist. Gleichzeitig ist die Abtastzeit nicht zu klein eingestellt, sodass nicht zu viele Abtastpunkte vermessen werden müssen. Vorteil der Einstellregel für die Abtastzeit basierend auf dem Signal-Rausch-Verhältnis

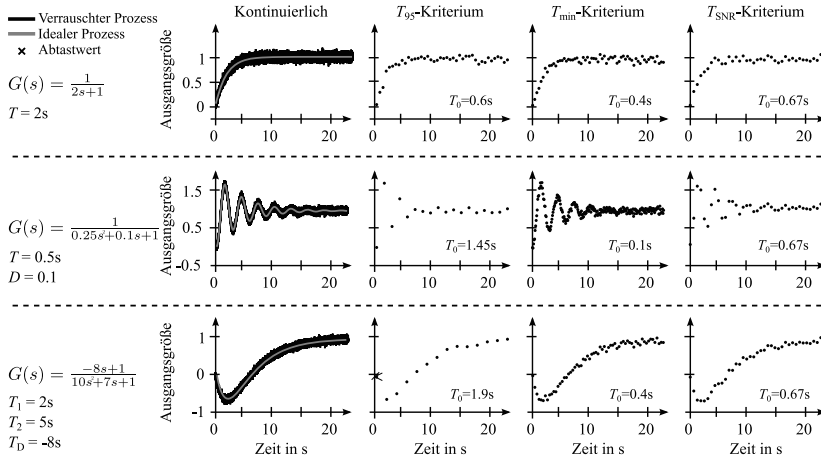


Bild E.2: Vergleich der kontinuierlichen und diskreten Sprungantworten verschiedener Prozesse $G(s)$ für das T_{95} -, T_{\min} - sowie das SNR-Einstellkriterium für die Abtastzeit.

ist, dass sie für lineare und nichtlineare Prozesse, unabhängig von den Prozesszeitkonstanten T_{95} bzw. T_{\min} , geeignet ist. Zudem kann die Abtastzeit automatisiert aus einem kurzen Messdatensatz bestimmt werden, welcher eine verrauschte Übergangsfunktion der Ausgangsgröße enthält.

Die Filterung auf die angepasste Abtastfrequenz wird unmittelbar im Echtzeitsystem realisiert, indem die Triggerung der Datenaufzeichnung für die jeweils modellierte Ausgangsgröße umgesetzt wird. Der sich ergebende Vorteil ist die unmittelbare Aufzeichnung der Messdaten in der benötigten Abtastfrequenz.

Durch die Integration von automatischen Datenaufbereitungsmethoden in die Online-Methodik wird eine Reduktion der Vermessungszeit erreicht, da die Daten direkt nach Abschluss der Messung aufbereitet vorliegen und ohne Benutzereingriff zur sofortigen Modellbildung genutzt werden können.

F Datenvorbereitung

Um Stell- und Ausgangsgrößen mit verschiedenen Größenordnungen (z.B. $0 \frac{\text{N}}{\text{m}^2} \leq p_{\text{Rail}} \leq 200.000.000 \frac{\text{N}}{\text{m}^2}$ und $0 \% \leq u_{\text{agr}} \leq 100 \%$) im Rahmen der experimentellen Modellbildung und automatisierten Online-Vermessung nutzen zu können, müssen diese vereinheitlicht werden. Dabei gibt es verschiedene Arten der Normierung. Bei der *Umskalierung* wird die um den minimalen Wert korrigierte Größe auf ihren maximalen Bereich bezogen

$$x_{i,\text{skal}} = \frac{x_i - \min(x)}{\max(x) - \min(x)}, \quad (\text{F.1})$$

wodurch eine Normierung auf den Einheitsbereich $[0, 1]$ erfolgt. Die Umskalierung wird im Rahmen der Online-Methode sowohl bei der Versuchsplanung (Kapitel 4) als auch bei der Variationsraumbestimmung (Kapitel 3) eingesetzt, damit die Größenordnung der verschiedenen Stellgrößen keinen Einfluss auf die Messpläne bzw. die Grenzraumregelung haben.

Die *Standardisierung* der Größe entspricht dem Bezug der um den Erwartungswert korrigierten Größe auf ihre Varianz

$$x_{i,\text{stand}} = \frac{x_i - \mu(x)}{\sigma(x)}, \quad (\text{F.2})$$

wodurch die Größe einen Erwartungswert von $\mu(x) = 0$ und eine konstante Einheitsvarianz von $\sigma(x) = 1$ erhält. Standardisierte Residuen e werden bei der Schätzung angenommen, damit die LS-Schätzung zu einer minimalvarianten linearen erwartungstreuen Schätzung (engl. BLUE, best linear unbiased estimator) wird. Dadurch wird die LS-Schätzung zur Schätzung mit der kleinsten Varianz aller linearen Schätzer [100].

Ist die Verteilung der Größe x nicht bekannt, wird anstelle der Standardisierung die *Studentisierung* durchgeführt, bei der anstatt des Erwartungswertes der arithmetische Mittelwert und anstatt der Varianz die Stichprobenvarianz verwendet werden:

$$x_{i,\text{stud}} = \frac{x_i - \bar{x}}{\sqrt{\frac{1}{N} \sum_k (x_k - \bar{x})^2}} \quad (\text{F.3})$$

Eine weitere Möglichkeit besteht in dem Bezug der Größe auf die Einheitslänge, indem sie durch die euklidische Länge des Vektors geteilt wird:

$$x_{i,\text{euklid}} = \frac{x_i}{\|x\|}. \quad (\text{F.4})$$

G Erweiterungen der LS-Parameterschätzung

Die in Abschnitt 5.1.1 beschriebenen LS- und RLS-Verfahren zur Parameterschätzung sind die grundlegenden Schätzverfahren. Im folgenden werden verschiedenen Erweiterungen mit den jeweiligen Eigenschaften vorgestellt, welche im Rahmen einer Parameterschätzung alternativ auch zum Einsatz kommen können.

Das *weighted* LS-Verfahren (WLS, [2]) ist durch eine individuelle Gewichtung der einzelnen Residuen bzw. Messungen auch für Fehler mit unterschiedlicher Varianz (Heteroskedastizität) geeignet. Die zu minimierende Zielfunktion beinhaltet somit zusätzlich die Gewichtungsmatrix W , welche eine Diagonalmatrix mit den jeweiligen Gewichtungen ist:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left(\mathbf{e}^T(\mathbf{x}, \hat{\boldsymbol{\theta}}) \mathbf{W} \mathbf{e}(\mathbf{x}, \hat{\boldsymbol{\theta}}) \right) \quad (\text{G.1})$$

Das *orthogonal* LS-Verfahren (OLS, [28]) minimiert nicht, wie das klassische LS-Verfahren, den Abstand in y -Richtung (siehe Gleichung (5.6)), sondern den senkrechten Abstand. Die zu minimierende Zielfunktion entspricht dabei der Summe der quadrierten Abstände in x - und y -Richtung und folgt zu

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}, \Psi_i^*} \left(\sum_{i=1}^N \left[\left(y_i(\mathbf{x}_i) - \Psi_i^*(\mathbf{x}_i) \hat{\boldsymbol{\theta}} \right)^2 + (\Psi_i - \Psi_i^*)^2 \right] \right) \quad (\text{G.2})$$

Hierbei sind die einzelnen Elemente $x_{i,j}^*$ in der Matrix Ψ_i^* die zu dem jeweiligen Regressor $x_{i,j}$ gehörenden Fußpunkte.

Sind die Regressoren mit Fehlern behaftet, kommt das *total* LS-Verfahren (TLS, [49, 102]) zum Einsatz. Dabei beinhaltet die Zielfunktion die Varianz-Kovarianz-Matrix \mathbf{M}_x der Regressoren \mathbf{x} sowie entsprechend \mathbf{M}_y des Regressanden \mathbf{y} :

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \left(\mathbf{e}_x^T(\mathbf{x}, \hat{\boldsymbol{\theta}}) \mathbf{M}_x^{-1} \mathbf{e}_x(\mathbf{x}, \hat{\boldsymbol{\theta}}) + \mathbf{e}_y^T(\mathbf{x}, \hat{\boldsymbol{\theta}}) \mathbf{M}_y^{-1} \mathbf{e}_y(\mathbf{x}, \hat{\boldsymbol{\theta}}) \right) \quad (\text{G.3})$$

Hierbei stellen \mathbf{e}_x bzw. \mathbf{e}_y die Residuen in x - bzw. y -Richtung dar, wie bei dem OLS-Verfahren.

Weitere Verfahren sind z.B. das *constrained* LS-Verfahren (CLS, [48]), welches Beschränkungen für die zu bestimmenden Parameter vorgibt, das *partial* LS-Verfahren (PLS, [169]), welches bei korrelierten Regressoren (Multikollinearität) zum Einsatz kommt sowie das *generalized* LS-Verfahren (GLS, [2]), welches eine Verallgemeinerung des WLS-Verfahrens ist, indem die Gewichtungsmatrix nicht auf eine Diagonalmatrix beschränkt wird. Dadurch können sowohl ungleiche Varianzen der Residuen (diagonale Elemente) sowie eine Korrelation derer (nichtdiagonale Elemente) einbezogen werden.

Sind die Modelle nicht-linear in ihren Parametern, kann nicht mehr eine geschlossene Lösung der Optimierung bestimmt werden. Stattdessen müssen entsprechende Optimierungsverfahren

wie z.B. der Levenberg–Marquardt Algorithmus [98] eingesetzt werden, um das Minimum der Zielfunktion zu bestimmen. Bei diesen *nonlinear* LS-Verfahren ist die Bestimmung der Parameter somit aufwendiger, weshalb sie für den Einsatz in einer Online-Methodik nicht gut geeignet sind.

H Vermessung von Kennfeldmodellen

Kennfeldmodelle

Kennfelder sind die einfachste Möglichkeit, einen nichtlinearen statischen Zusammenhang abzubilden und werden aufgrund ihres einfachen Aufbaus und der schnellen Berechnung oft eingesetzt. Jede Kennfeldstützstelle besteht in diesem Fall aus einer Position \mathbf{c}_i im Wertebereich der Modelleingangsgrößen $\mathbf{x}_{\min} \leq \mathbf{c}_i \leq \mathbf{x}_{\max}$ und einer Höhe $w_i = \hat{y}_i(\mathbf{c}_i)$.

Der simulierte Ausgang des Kennfeld-Modells \hat{y} folgt für beliebige Modelleingangsgrößen \mathbf{x}_i innerhalb des Bereichs $[\mathbf{x}_{\min}, \mathbf{x}_{\max}]$ zu

$$\hat{y}(\mathbf{x}_i, \mathbf{c}) = \sum_{p=1}^{2^P} w_{(2^P+1-p)} \cdot \Phi_p(\mathbf{x}_i, \mathbf{c}(\mathbf{x}_i)) \quad (\text{H.1})$$

Der Modellausgang ergibt sich mittels einer P -dimensionalen Interpolation der Positionen und Höhen aller 2^P umliegenden Stützstellen [68]. Algorithmen zur Bestimmung der umliegenden Stützstellen können aus [136] entnommen werden, weshalb auf sie im Folgenden nicht eingegangen wird. Jede der umliegenden Stützstellen hat dabei einen linearen Einfluss ihrer Höhe auf den Modellausgang, welcher mit der jeweiligen Gewichtungsfunktion $\Phi_p(\mathbf{x}_i, \mathbf{c}(\mathbf{x}_i))$ abgebildet wird. In Bild H.1 ist dieser Zusammenhang für den zweidimensionalen Fall dargestellt [146].

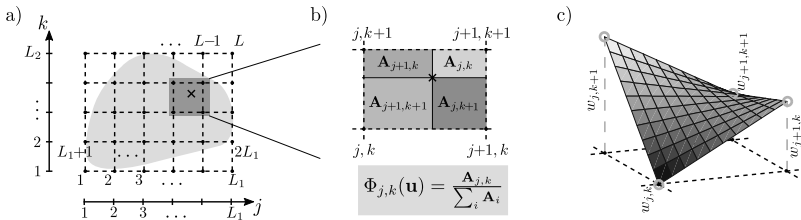


Bild H.1: a) Äquidistante Verteilung von Kennfeldstützstellen für ein Kennfeldmodell, welches den gesamten Eingangsgrößenbereich (hellgrau) abdeckt. b) Darstellung der 2-dimensionalen linearen Flächeninterpolation zur Bestimmung der Gewichtungsfunktion. c) Resultierender Modellausgang für alle Modelleingangsgrößen innerhalb eines Kennfeldelements mit den 4 Stützstellen $[[c_{1,1}, c_{2,1}, w_1], \dots, [c_{1,4}, c_{2,4}, w_4]]^T$ [146].

Der Modellausgang \hat{y} folgt dabei für einen Modelleingang $\mathbf{x}_i = (x_{1,i}, x_{2,i})$ und den 4 umliegenden Stützstellen $\mathbf{c}(\mathbf{x}_i) = ([c_{1,1}(x_{1,i}), c_{2,1}(x_{2,i})], \dots, [c_{1,4}(x_{1,i}), c_{2,4}(x_{2,i})])^T$ zu

$$\hat{y}(\mathbf{x}_i, \mathbf{c}) = \sum_{p=1}^4 w_{(5-p)} \cdot \Phi_p(x_{1,i}, x_{2,i}, c_{1,1}, c_{2,1}, \dots, c_{1,4}, c_{2,4}) \quad (\text{H.2})$$

Die Gewichtungsfunktion Φ_p kann mithilfe der linearen (Flächen-) Interpolation aus dem Modelleingang \mathbf{x}_i und den 4 umliegenden Stützstellen $\mathbf{c}(\mathbf{x}_i)$ berechnet werden, indem jeweils die

durch eine Stützstelle und den Modelleingang aufgespannte Teilfläche A_p sowie die durch die 4 Stützstellen aufgespannte Gesamtfläche bestimmt wird:

$$\Phi_p(x_{1,i}, x_{2,i}) = \frac{A_p}{|c_{1,2} - c_{1,1}| |c_{2,3} - c_{2,1}|}$$

$$\text{mit } A_p = \begin{cases} |c_{1,1} - x_{1,i}| |c_{2,1} - x_{2,i}| & \text{für } p=1 \text{ (links unten)} \\ |c_{1,2} - x_{1,i}| |c_{2,2} - x_{2,i}| & \text{für } p=2 \text{ (rechts unten)} \\ |c_{1,3} - x_{1,i}| |c_{2,3} - x_{2,i}| & \text{für } p=3 \text{ (links oben)} \\ |c_{1,4} - x_{1,i}| |c_{2,4} - x_{2,i}| & \text{für } p=4 \text{ (rechts oben)} \end{cases} \quad (\text{H.3})$$

Zur Identifikation eines Kennfeldmodells aus einem Messdatensatz müssen die Positionen und die Höhen aller Stützstellen bestimmt werden. Dabei ist zu berücksichtigen, dass der Modellausgang \hat{y} sich linear mit der Höhe und nichtlinear mit der Position ändert, aufgrund der Interpolation des Modellausgangs aus den umliegenden Stützstellen (siehe (H.1)).

Die Positionen aller Stützstellen c_i werden entweder basierend auf Vorwissen (z.B. an die Prozessnichtlinearität angepasst) [19] oder optimiert im Versuchsraum festgelegt [40], wobei für letzteres der gesamte Messdatensatz vorliegen und eine entsprechende Größe aufweisen muss. Dabei wird iterativ ein geschachtelter Algorithmus eingesetzt, welcher in einer äußeren Schleife die Stützstellenpositionen mit einem nichtlinearen Optimierer (Trust-Region-Verfahren [25]) bestimmt. Anschließend werden in einer inneren Schleife die Stützstellenhöhen für die neuen Positionen angepasst. Das Vorgehen wird iterativ wiederholt, bis die Verlustfunktion (z.B. Fehlerquadratsumme) konvergiert.

Die 2^P zugehörigen Stützstellenhöhen ergeben sich dabei entweder direkt aus dem Messdatensatz, falls bei der jeweiligen Stützstellenposition ein gemessener Wert vorliegt, oder durch eine LS-Schätzung der Höhen als Parameter [163]. In diesem Fall kann eine LS-Schätzung, mit den entsprechenden Annahmen für die Residuen nach (5.7), durchgeführt werden mit der Regressormatrix

$$\Psi = \begin{bmatrix} \Phi_1(x_1) & \Phi_2(x_1) & \cdots & \Phi_V(x_1) \\ \Phi_1(x_2) & \Phi_2(x_2) & \cdots & \Phi_V(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \Phi_1(x_N) & \Phi_2(x_N) & \cdots & \Phi_V(x_N) \end{bmatrix}. \quad (\text{H.4})$$

sowie dem zu schätzenden Parametervektor, der die Stützstellenhöhen enthält,

$$\Theta^T = (w_1, w_2, \dots, w_V). \quad (\text{H.5})$$

Die einzelnen Gewichtungsfunktionen der Regressormatrix $\Phi_j(x_i)$ bestimmen dabei den Einfluss des jeweiligen Messpunktes x_i (mit $i = 1, \dots, N$) auf die jeweilige Stützstelle c_j (mit $j = 1, \dots, V$). Die Höhen können dann nach Gleichung (5.7) zu

$$\hat{\Theta} = (\Psi^T \Psi)^{-1} \Psi^T y. \quad (\text{H.6})$$

bestimmt werden. Sind nicht alle N Messdaten zu Beginn der Schätzung vorhanden, kann auch hier das RLS-Verfahren zum Einsatz kommen. Dabei werden die Stützstellenhöhen des dis-

kreten Zeitpunktes k mit jeder neuen Messung zu den Zeitpunkten $k + 1$ entsprechend Gleichung (5.8) mit einem Gewichtungsvektor $\gamma(k + 1)$ aktualisiert:

$$\hat{\Theta}(k + 1) = \hat{\Theta}(k) + \gamma(k + 1) (\mathbf{y}(k + 1) - \hat{\mathbf{y}}(k + 1)). \quad (\text{H.7})$$

Das Vorgehen zur Bestimmung geeigneter Gewichtungen $\gamma(k + 1)$ sowie der Startwerte $\gamma(0)$ und $\hat{\Theta}(0)$ bei der rekursiven Bestimmung von Kennfeldmodellen ist detailliert in [146] beschrieben, weshalb hier nicht darauf eingegangen wird. Weiterhin kann die rekursive Berechnung der Stützstellenhöhen durch geeignete Algorithmen gemäß [166] beschleunigt werden.

Für Modelleingangsgrößen \mathbf{x}_i außerhalb des Bereichs $[\mathbf{x}_{\min}, \mathbf{x}_{\max}]$ ist der Modellausgang nicht definiert und wird somit auf den Wert am Rand des Stellbereichs festgelegt.

Durch den exponentiellen Anstieg der Stützstellen mit der Anzahl an Eingangsgrößen, und somit auch dem exponentiellen Anstieg der zu berechnenden Gewichtungsfunktionen, steigt der Rechenaufwand entsprechend. Kennfelder werden deshalb vor allem für zweidimensionale Problemstellungen verwendet. Für höherdimensionale Probleme (bis $P = 5$) können hierarchische Kennfeldmodelle eingesetzt werden [162]. Dabei wird ein exponentieller Anstieg der Stützstellen vermieden, indem die Stützstellen in die relevanten Bereiche des gesamten Stellbereichs gelegt werden. Der erhöhte Rechenaufwand zur Berechnung der Gewichtungsfunktionen bleibt aber erhalten. Eine Übersicht zu Kennfeldmodellen ist in [147] zu finden.

Eine weitere Alternative stellt die additive oder multiplikative Überlagerung oder Verschachtelung von ein- und zweidimensionalen Kennfeldmodellen dar [146]. Dabei muss ein erhöhter Aufwand zur Versuchsplanung bzw. Vermessung aufgebracht werden, um für jedes einzelne Kennfeld \hat{f}_i den optimalen Messplan zu bestimmen. Bei der Vermessung und Modellierung eines unbekannten Prozesses sind Modelleingangsgrößen unbekannt, weshalb verschachtelte Kennfelder hierfür nicht geeignet sind. Dem gegenüber steht jedoch der Vorteil, im Rahmen einer semiphysikalischen Modellbildung sehr gut an einen Prozess anpassbar zu sein. Dabei müssten jedoch die Modelleingangsgrößen sowie die Art ihrer Überlagerung bzw. Verschachtelung bekannt sein.

Kennfeldglättung

Bei der Schätzung sollte, um eine Überanpassung des Kennfeldmodells an die Messdaten [19] bzw. die ungenaue Interpolation in Bereichen mit wenig bzw. keinen Messpunkten [146] zu vermeiden, zusätzlich die Glätte des Kennfelds beachtet werden. Generell werden glatte Kennfelder bevorzugt, um die physikalische Interpretierbarkeit und somit einen direkten Einsatz in der Motorapplikation sicherzustellen und eine Beseitigung von Störsignaleinflüssen zu erreichen.

Dazu wird die Regressormatrix um einen Strafterm erweitert, welcher die Flexibilität des Kennfeldmodells einschränkt:

$$\Psi_+ = \begin{bmatrix} \Psi \\ \Omega \end{bmatrix} \quad (\text{H.8})$$

Um ein glattes Kennfeld zu schätzen, muss als Bestrafung die Krümmung der inneren Kennfeldstützstellen, welche jeweils vier umliegende Stützstellen haben, sowie der Randstützstellen, welche jeweils drei umliegende Stützstellen haben, zu null gesetzt werden. Die Anzahl W an inneren Stützstellen ergibt sich für ein äquidistantes Kennfeldrastrer mit L_p -Stufen in der jeweiligen Dimension $p = 1, \dots, P$ zu

$$W_I = \prod_{p=1}^P (L_p - 2) \quad (\text{H.9})$$

sowie die Anzahl an Randstützstellen zu

$$W_R = \sum_{p=1}^P (L_p 2^{P-1}). \quad (\text{H.10})$$

Für die 2^P -Eckstützstellen kann keine Krümmung bestimmt werden, da Eckpunkte keine zwei umliegenden Stützstellen entlang einer Dimension haben. Für den zweidimensionalen Fall folgen die Differenzenquotienten zweiter Ordnung, welche der zweiten Ableitung bzw. Krümmung an der Stützstelle entsprechen, zu

$$\frac{\partial^2 w_i}{\partial u_1^2} = \frac{w_{i-1} - 2w_i + w_{i+1}}{\Delta u_1^2} \quad (\text{H.11})$$

$$\frac{\partial^2 w_i}{\partial u_2^2} = \frac{w_{i-L_1} - 2w_i + w_{i+L_1}}{\Delta u_2^2}. \quad (\text{H.12})$$

Die Krümmungsmatrix Γ ergibt sich zu einer Matrix mit $(P \cdot W_I + W_R)$ -Zeilen und V -Spalten. Für jede innere Stützstelle sind dabei P -Zeilen und für jede Randstützstelle eine Zeile vorhanden, die jeweils die Krümmung in einer Dimension beschreiben. Im zweidimensionalen Fall folgt somit:

$$\Gamma = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 1 & -2 & 1 & 0 & \dots & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & 1 & 0 & \dots & \dots & -2 & 0 & \dots & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots & 1 & -2 & 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & 1 & 0 & \dots & \dots & -2 & 0 & \dots & \dots & 1 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & 1 & -2 & 1 \end{bmatrix} \begin{matrix} \rightarrow c_{1,2} \\ \rightarrow c_{1,3} \\ \rightarrow \vdots \\ \rightarrow c_{1,L_1-1} \\ \rightarrow c_{2,L_1+1} \\ \rightarrow c_{1,L_1+2} \\ \rightarrow c_{2,L_1+2} \\ \rightarrow \vdots \\ \rightarrow c_{1,V-1} \end{matrix} \quad (\text{H.13})$$

Der von der Messpunktanzahl unabhängige Strafterm Ω folgt nach [146, 125] mit der Krümmungsmatrix Γ und dem Gewichtungsfaktor λ zu

$$\Omega = \lambda \Gamma \sqrt{\frac{\text{trace}(\Psi^T \Psi)}{\text{trace}(\Gamma^T \Gamma)}} \quad (\text{H.14})$$

Dabei ist λ ein Zeilenvektor mit $(P \cdot W_I + W_R)$ -Elementen, welche den Gewichtungsfaktoren $\lambda_p \in [0, \infty]$ für jede Dimension $p = 1, \dots, P$ entsprechen:

$$\lambda = (\lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_P \quad \lambda_1 \quad \lambda_2 \quad \dots \quad \lambda_P \quad \lambda_1 \quad \lambda_2 \quad \dots) \quad (\text{H.15})$$

Je größer λ_p gewählt wird, desto stärker wird entlang dieser Dimension geglättet. Entsprechend folgt für $\lambda_p = 0$ keine Regularisierung des Kennfeldmodells. Die Bestimmung des Kennfeldmodells bleibt mit dieser Erweiterung eine lineare Schätzung der Stützstellenhöhen, wodurch die Höhen aus der LS-Schätzgleichung folgen:

$$\hat{\boldsymbol{\theta}}_+ = (\boldsymbol{\Psi}_+^T \boldsymbol{\Psi}_+)^{-1} \boldsymbol{\Psi}_+^T \mathbf{y}_+ = (\boldsymbol{\Psi}^T \boldsymbol{\Psi} + \boldsymbol{\Omega}^T \boldsymbol{\Omega})^{-1} \boldsymbol{\Psi}^T \mathbf{y}. \quad (\text{H.16})$$

Bei gleichbleibenden Positionen der Stützstellen \mathbf{u} ist die Krümmungsmatrix Γ konstant. Die geglätteten Kennfeldhöhen $\hat{\mathbf{y}}_{\text{glatt}}$ ergeben sich somit direkt aus den ursprünglichen Kennfeldhöhen $\hat{\mathbf{y}}$ zu

$$\hat{\mathbf{y}}_{\text{glatt}} = (\mathbf{I} + \boldsymbol{\Omega}^T \boldsymbol{\Omega})^{-1} \mathbf{I}^T \hat{\mathbf{y}} \quad (\text{H.17})$$

Der Vorteil einer solchen Darstellung ist, dass die Inverse nur einmalig berechnet werden muss, da diese nur von den konstanten $\boldsymbol{\Omega}$ und \mathbf{I} abhängt.

Eine solche Regularisierung entspricht der Reduktion der effektiven Parameter in dem Kennfeldmodell, wodurch ein geglättetes Kennfeld entsteht. Effektive Parameter entsprechen in einem Kennfeldmodell den Stützstellenhöhen, welche einen dominanten Einfluss auf den Modelausgang haben [111].

I Identifikation von richtungsabhängigen Prozessen

Alle im Abschnitt 5.1.2 beschriebenen Verfahren zur Identifikation von Modellen bestimmen bei der Parameterschätzung mithilfe einer Zielfunktion die im Mittel besten Parameter. Dieses Vorgehen ist für alle Prozesse geeignet, bei denen keine Richtungsabhängigkeit vorhanden ist.

Im Falle von Prozessen, bei denen der Prozessausgang eine Richtungsabhängigkeit aufweist, führt die klassische Parameterschätzung zu einer Mittlung der Richtungsabhängigkeit, sodass der Prozessausgang in keine Richtung genau modelliert wird. Ursache für die Richtungsabhängigkeit ist entweder eine Hysterese im System (z.B. Aufgrund des Einsatzes eines Magneten zur Verstellung, wie bei dem Zusammenhang zwischen PWM-Ansteuerung und Position bei einem elektrischen AGR-Magnetventil [122]) oder die Kombination mehrerer Aktorprinzipien je nach Stellrichtung innerhalb eines Aktors (z.B. Kombination einer mechanischen Feder mit einem pneumatischen Zusammendrücken dieser, wie bei dem Zusammenhang zwischen PWM-Ansteuerung und Position beim pneumatischer Aktor des VTG-Laders [143]). Tan hat in [155] eine Zusammenstellung verschiedener Systeme mit Richtungsabhängigkeit in den verschiedensten technischen Bereichen erstellt, wie z.B. bei Gasturbinen, welche einen Zusammenhang zwischen Kraftstoffmasse und Verdichterdrehzahl bzw. Temperatur aufweisen [45].

Die Identifikation von richtungsabhängigen dynamischen Prozessen wurde erstmals von Godfrey und Briggs betrachtet [44] und von Godfrey und Moore auf Gasturbinen angewandt [45]. Sie identifizierten richtungsabhängige Modelle für Prozesse mit Dynamiken 1. Ordnung mithilfe der Kreuzkorrelation. Barker setzt zur richtungsabhängigen Modellbildung Wiener-Modelle mit 4 Pfaden ein. Hierbei nutzt er jeweils einen für den konstanten, linearen, quadratischen und kubischen Term [10]. Die Parameter werden auch wieder mit der Kreuzkorrelation bestimmt. Diese Art der Modellierung eignet sich für richtungsabhängige Systeme mit einer dynamischen Ordnung von 1 sowie einer Eingangsgröße. Die Identifikation von richtungsabhängigen Systemen mit einer höheren dynamischen Ordnung erfolgt mit PWL-Modellen (engl. *piecewise linear* [15]). Dabei werden lineare Teilmodelle für jede Richtung identifiziert und bei der Modellsimulation zwischen den Teilmodellen je nach Richtung hin- und her geschaltet. Rosenqvist nutzt beide Arten der Modellbildung um den Gassensor in einer elektronischen Nase zu modellieren [132]. Eine weitere Möglichkeit besteht im Einsatz von bilinearen Zustandsraummodel-

len zur richtungsabhängigen Modellierung, welche sich für ein SISO-Prozess aus

$$\begin{aligned}
 \dot{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{b}u + \mathbf{u}\mathbf{D}\mathbf{x} \\
 &= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \\ -\alpha_0 & \cdots & \cdots & \cdots & -\alpha_{C-1} \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ \vdots \\ \vdots \\ 0 \\ 1 \end{bmatrix} u + u \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \\ \rho_0 & \cdots & \rho_{C-1} \end{bmatrix} \mathbf{x} \quad (I.1) \\
 y &= \mathbf{c}^T \mathbf{x} \\
 &= [\beta_0 \quad \cdots \quad \beta_B \quad 0 \quad \cdots \quad 0] \mathbf{x}
 \end{aligned}$$

ergeben [30, 52]. Dabei sind $u \in R$ die Stellgröße, $y \in R$ die Ausgangsgröße, $\mathbf{x} \in R^n$ der Zustandsvektor sowie C die dynamische Ordnung des Systems. Die Richtungsabhängigkeit ergibt sich durch die Parameter α_i , β_i und ρ_i , welche für ein richtungsabhängiges System 1. Ordnung (und somit $C = 1$), welches eine Verstärkung K_{-1} und Zeitkonstante T_{-1} in die eine Richtung sowie K_{+1} und T_{+1} in die andere Richtung hat, nach [170] zu

$$\begin{aligned}
 \alpha_0 &= \frac{1}{2T_{-1}} + \frac{1}{2T_{+1}} \\
 \beta_0 &= \frac{K_{-1} + K_{+1}}{T_{-1} + T_{+1}} \\
 \rho_0 &= \frac{1}{2VT_{+1}} - \frac{1}{2VT_{-1}}
 \end{aligned} \quad (I.2)$$

folgen. Hierbei ist V die Amplitude des binären Anregungssignals. Zusätzlich ist zu berücksichtigen, dass der Offset

$$b_0 = \frac{V(K_{+1}T_{-1} - K_{-1}T_{+1})}{T_{-1} + T_{+1}} \quad (I.3)$$

vom Modellausgang subtrahiert werden muss [170]. Auch rekurrente neuronale Netze (neuronale Netze mit direkten, indirekten und seitlichen Rückkopplungen) eignen sich, richtungsabhängige Systeme mit einer Dynamik 1. Ordnung zu modellieren [164, 156], sind jedoch sehr zeitaufwendig und schwierig zu trainieren. Eine weitere Möglichkeit zur Identifikation von richtungsabhängigen Prozessen ist die Berücksichtigung der Änderung der Modelleingangsgröße als z-Regressor bei LOLIMOT-Netzmodellen [120].

Die bisher beschriebenen Verfahren haben den Nachteil, entweder stark eingeschränkt (linear, mit einer Eingangsgröße, mit einer Dynamik 1. Ordnung) oder aufwendig identifizierbar zu sein. Um die Anwendung der Online-Methodik für die Identifikation von richtungsabhängigen Prozessen zu ermöglichen, wird ein alternatives Verfahren entwickelt, welches sich zum Einsatz im Rahmen dieser eignet. Hierfür wird eine Schaltfunktion $\sigma(k) = \{-1, 1\}$ gemäß [155] eingeführt, für die gilt:

$$\sigma(k) = \begin{cases} 1 & \text{für } u(k) - u(k-1) > 0 \\ -1 & \text{für } u(k) - u(k-1) < 0 \\ \sigma(k-1) & \text{für } u(k) - u(k-1) = 0. \end{cases} \quad (I.4)$$

Eine Formulierung in Abhängigkeit von der Ausgangsgröße $\hat{y}(k)$ wäre auch möglich, würde aber den Einsatz auf nicht schwingungsfähige Prozesse mit monotonen Ausgangsgrößen beschränken. Aus der Schaltfunktion kann die Regressorgewichtungsfunktion Σ gebildet werden

$$\Sigma = \left[\frac{1+\sigma(k)}{2} \quad \frac{1-\sigma(k)}{2} \right], \quad (I.5)$$

welche je nach Richtung als Element eine 1 oder 0 enthält. Der gewichtete Regressorvektor $\Psi_{\Delta}(k)$ entsteht, indem das Kronecker-Produkt von Σ mit dem um den konstanten Offset-Regressor reduzierten Regressorvektor $\Psi_{-}(k)$ (entspricht der Zeile k der Regressormatrix aus Gleichung (5.2), jedoch ohne der 1. Spalte) gebildet und anschließend der konstante Regressor wieder hinzugefügt wird:

$$\Psi_{\Delta}(k) = \begin{bmatrix} 1 & \Sigma \otimes \Psi_{-}(k) \end{bmatrix} \quad (I.6)$$

Aus diesem kann für jeden der $k = 1, \dots, N$ Messpunkte die gewichtete Regressormatrix Ψ_{Δ} aufgestellt werden:

$$\Psi_{\Delta} = \begin{bmatrix} 1 & \frac{1+\sigma(1)}{2}x_{1,1} & \cdots & \frac{1+\sigma(1)}{2}x_{R-1,1} & \frac{1-\sigma(1)}{2}x_{1,1} & \cdots & \frac{1-\sigma(1)}{2}x_{R-1,1} \\ 1 & \frac{1+\sigma(2)}{2}x_{1,2} & \cdots & \frac{1+\sigma(2)}{2}x_{R-1,2} & \frac{1-\sigma(2)}{2}x_{1,2} & \cdots & \frac{1-\sigma(2)}{2}x_{R-1,2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \frac{1+\sigma(N)}{2}x_{1,N} & \cdots & \frac{1+\sigma(N)}{2}x_{R-1,N} & \frac{1-\sigma(N)}{2}x_{1,N} & \cdots & \frac{1-\sigma(N)}{2}x_{R-1,N} \end{bmatrix} \quad (I.7)$$

Im Gegensatz zu dem WLS-Verfahren wird hierbei nicht jede einzelne Messung, sondern jeder Regressor in jeder Messung gewichtet. Die Gewichtung erfolgt mit den Faktoren 1 bzw. 0, je nachdem ob die Schaltfunktion 1 oder -1 ist. Hierdurch können in Abhängigkeit der aktuellen Richtung getrennte Parameter für die jeweiligen Regressoren bestimmt werden. Die Parameter werden dabei mit der LS-Schätzgleichung bestimmt

$$\hat{\Theta} = (\Psi_{\Delta}^T \Psi_{\Delta})^{-1} \Psi_{\Delta}^T y \quad (I.8)$$

Dem Vorteil dieses Verfahrens, die Richtung identifizieren zu können, steht der Nachteil gegenüber, dass die Anzahl der Modellparameter von R auf $R + R - 1$ erhöht wird.

In Bild I.1 sind sowohl (a) die Modellbildung des statischen Zusammenhangs zwischen der Ansteuerspannung u_{VTG} und der Position s_{VTG} eines pneumatischen VTG-Aktors mit einem richtungsabhängigen LOPOMOT mit 2 Teilungen als auch (b) die Modellierung eines richtungsabhängigen dynamischen Beispiel-Prozesses mit einem richtungsabhängigen Polynom 1. statischer bzw. 2. dynamischer Ordnung in OE-Anordnung dargestellt. Der simulierte Prozess hat dabei die Übertragungsfunktion

$$G(s) = \begin{cases} \frac{1}{3s^2 + s + 1} & \text{für } u(k) - u(k-1) > 0 \\ \frac{0.5}{4s^2 + 1.5s + 0.5} & \text{für } u(k) - u(k-1) < 0. \end{cases} \quad (I.9)$$

Die Richtungsabhängigkeit im pneumatischen VTG-Aktor entsteht durch den unterschiedlich schnellen Druckaufbau und Druckabbau in die zwei Verstellrichtungen [143]. Wie in Bild I.1 dargestellt, lassen sich beide Richtungsabhängigkeiten mit dem vorgestellten Verfahren mit einer hohen Qualität ($\text{RSE}_{\text{BSP-Prozess}} = 0.303$ bzw. $\text{RSE}_{\text{VTG-Aktor}} = 0.069$) modellieren, während

eine nicht richtungsabhängige Schätzung zu einem fehlerhaften Modell führt ($\text{RSE}_{\text{BSP-Prozess}} = 1.152$ bzw. $\text{RSE}_{\text{VTG-Aktor}} = 1.833$). Im Fall des pneumatischen VTG-Aktors lässt sich ein solches richtungsabhängiges Modell im Rahmen einer Positionsregelung einsetzen, indem dieses als statische Vorsteuerung der Stellgröße aufgeschaltet wird.

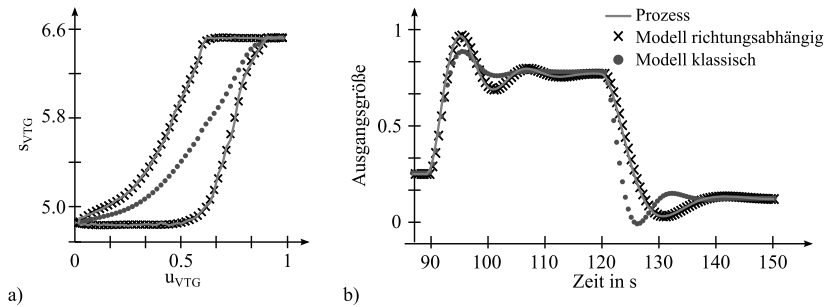


Bild I.1: Vergleich der gemessenen Ausgangsgröße mit jeweils einer klassisch modellierten und richtungsabhängig modellierten Modellausgangsgröße eines a) vermessenen pneumatischen VTG-Aktors sowie b) eines simulierten richtungsabhängigen Beispielsprozesses.

Literaturverzeichnis

- [1] ADAC: *Jährliche Durchschnittspreise Kraftstoffe seit 1950*.
<https://www.adac.de/infotestrat/tanken-kraftstoffe-und-antrieb/kraftstoffpreise/kraftstoff-durchschnittspreise/>, Abruf: 31. Juli. 2016
- [2] AITKEN, A. C.: On Least Squares and Linear Combination of Observations. In: *Proceedings of the Royal Society of Edinburgh* 55 (1936), S. 42–48
- [3] ALEKSANDROV, M.: *Weiterentwicklung und Beurteilung von Vermessungsstrategien zur Variationsraumbestimmung im Hinblick auf den Einsatz modellgestützter Methoden in der Nfz-Motorenentwicklung*, TU Darmstadt, Diplomarbeit, 2006
- [4] AURENHAMMER, F.: Voronoi diagrams – a survey of a fundamental geometric data structure. In: *ACM Computing Surveys* 23 (1991), Nr. 3, S. 345–405
- [5] AVL: *Cameo Test Measure*. <https://www.avl.com/-/avl-cameo-test-measure>, Abruf: 31. Mai. 2015
- [6] BACKHAUS, K.; ERICHSON, B.; PLINKE, W.; WEIBER, R.: *Multivariate Analysemethoden*. 13. Auflage. Springer, Berlin, 2011
- [7] BANDEMER, H.; BELLMANN, A.: *Statistische Versuchsplanung*. 4. Auflage. Teubner, Stuttgart, 1994
- [8] BÄNFER, O.; NELLES, O.: Polynomial model tree (POLYMOT) - A new training algorithm for local model networks with higher degree polynomials. In: *2009 IEEE International Conference on Control and Automation*. Christchurch: IEEE, 2009, S. 1571–1576
- [9] BARBER, C. B.; DOBKIN, D. P.; HUHDANPAA, H.: The Quickhull Algorithm for Convex Hulls. In: *ACM Transactions on Mathematical Software* 22 (1996), S. 469–483
- [10] BARKER, H. A.; GODFREY, K. R.; TAN, A. H.: Optimised Wiener models for direction-dependent dynamic systems. In: *Proceedings of the American control conference*, 2001, S. 4880–4881
- [11] BAUMANN, W.; RÖPKE, K.; STELZER, S.; FRANK, A.: Model-Based Calibration Process for Diesel Engines. In: *4. Internationales Symposium für Entwicklungsmethodik*. Wiesbaden: AVL, 2011, S. 197–208
- [12] BAUMANN, W.; SCHAUM, S.; K., Röpke; KNAACK, M.: Excitation Signals for Nonlinear Dynamic Modeling of Combustion Engines. In: *17th IFAC World Congress*. Seoul, Korea: IFAC, Juli 2008, S. 1066–1067

- [13] BAZARSUREN, U.; KNAAK, M.; SCHAUM, S.; GÜHMANN, C.: Determining Adjustment Ranges for Model-based Approaches Using Support Vector Machines. In: *2006 IEEE International Conference on Control Applications*. München: IEEE, Oktober 2006, S. 2066–2071
- [14] BERGER, B.: *Modeling and Optimization for Stationary Base Engine Calibration*, TU München, Dissertation, 2012
- [15] BILLINGS, S. A.; VOON, W.S.F.: Piecewise linear identification of non-linear systems. In: *International Journal of Control* 46 (1988), Nr. 1, S. 215–235
- [16] BISHOP, C. M.: *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford, 1995
- [17] BITTERMANN, A.; KRANAWETTER, E.; KRENN, J.; LADEIN, B.; EBNER, T.; ALTENSTRASSER, H.; KOEGELER, H.-M.; GSCHWEITL, K.: Emissions Development of Vehicle Diesel Engines by Means of DoE and Computer Simulation. In: *MTZ worldwide Edition* 06 (2004), S. 15–19
- [18] BLESSING, P.: *Ein Verfahren zur Identifikation von linearen, stochastisch gestörten Mehrgößensystemen*. Projekt Prozeßlenkung mit DV-Anlagen, Band 181, Gesellschaft für Kernforschung mbH, Karlsruhe, 1980
- [19] BÖHME, D.: *Methoden zur effizienten Basisapplikation für Luftpfad- und Abgastemperaturmodelle von Ottomotoren*. Fortschritt-Berichte VDI: Nr. 767 in Reihe 12, VDI, Düsseldorf, 2012
- [20] BREDENBECK, J.: Statistische Versuchplanung für die Online-Optimierung von Verbrennungsmotoren. In: *MTZ - Motortechnische Zeitschrift* 60 (1999), Nr. 11, S. 740–744
- [21] BREIMAN, L.: Hinging Hyperplanes for Regression, Classification, and Functionapproximation. In: *IEEE Transactions on Information Theory* 39 (1993), Nr. 3, S. 999–1013
- [22] BRONSTEIN, I. N.; SEMENDJAJEW, K. A.; MUSIOL, G.; MUHLIG, H.: *Taschenbuch der Mathematik*. Verlag Harri Deutsch, 2000
- [23] BURNHAM, K. P.; ANDERSON, D. R.: *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer, Berlin, 2002
- [24] CHAUDOYE, F.; CASTAGNÉ, M.; SINOQUET, D.; WAHL, F.: Modelling engine operating space for DoE calibration methods. In: RÖPKE, K. (Hrsg.): *Design of Experiments (DoE) in Engine Development V*, Expert-Verlag, Renningen, 2009, S. 107–121

- [25] COLEMAN, T. F.; LI, Y.: On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds. In: *Mathematical Programming* 67 (1994), Nr. 1-3, S. 189–224
- [26] DAVISON, A. C.; HINKLEY, D. V.: *Bootstrap Methods and their Application*. Cambridge University Press, 1997
- [27] DEB, K.; PRATAP, A.; AGARWAL, S.; MEYARIVAN, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. In: *IEEE Transactions on Evolutionary Computation* 6 (2002), Nr. 2, S. 182–197
- [28] DEMING, W. E.: *Statistical Adjustment of Data*. Dover Publications, 1943
- [29] DOLT, R.: *Indizierung in der Motorentwicklung: Messtechnik, Datenauswertung, Anwendung und Kombination mit optischen Messverfahren*. Verlag Moderne Industrie, 2006
- [30] DUNOYER, A.; BALMER, L.; BURNHAM, K. J.; JAMES, D. J. G.: On the discretization of single-input single-output bilinear systems. In: *International Journal of Control* 68 (1997), Nr. 2, S. 361–372
- [31] EFRON, B.; HASTIE, I. T. and J. T. and Johnstone; TIBSHIRANI, R.: Least Angle Regression. In: *The Annals of Statistics* 32 (2004), Nr. 2, S. 1–67
- [32] ERIKSSON, L.; JOHANSSON, E.; KETTANEH-WOLD, N.; WIKSTRÖM, C.; WOLD, S.: *Design of Experiments - Principles and Applications*. Umetrics Academy, 2008
- [33] ETAS: *Messung, Steuergerätekalibrierung und Diagnose-Entwicklungslösungen für eingebettete Systeme im Fahrzeug*. http://www.etas.com/download-center-files/products_catalogs/MCD_Brochure_de_v2.pdf, Abruf: 30. April. 2016
- [34] EUROPÄISCHE UNION: *Richtlinie Nr. 70/220/EWG des Rates*. 1970
- [35] EUROPÄISCHE UNION: *Verordnung Nr. 715/2007/EG des Europäischen Parlaments und des Rates*. 2007
- [36] EYKHOFF, P.: *System Identification*. John Wiley & Sons, Hoboken, 1974
- [37] FAURE, H.: Good permutations for extreme discrepancy. In: *Journal of Number Theory* 42 (1992), S. 47–56
- [38] FEDOROV, V. V.: *Theory of optimal experiments*. Academic Press, 1972
- [39] FELLER, W.: *An Introduction to Probability Theory and Its Applications*. Bd. 1. John Wiley & Sons, Hoboken, 1968
- [40] FINK, A.: *Nonlinear Control Based on Local Linear Neuro-Fuzzy Models*. Fortschritt-Berichte VDI: Nr. 1096 in Reihe 8, VDI, Düsseldorf, 2006

- [41] GAUSS, C. F.; DAVIS, C. H.: *Theory of the motion of the heavenly bodies moving about the sun in conic sections: a translation of Gauss's "Theoria motus."*. Boston, Little, Brown and Company, 1857
- [42] GILL, P. E.; MURRAY, W.; WRIGHT, M. H.: *Practical Optimization*. Academic Press, London, 1986
- [43] GILMOUR, S. G.: The interpretation of Mallows' Cp statistic. In: *The Statistician* 45 (1996), Nr. 1, S. 49–56
- [44] GODFREY, K. R.; BRIGGS, P. A. N.: Identification of processes with direction-dependent dynamic responses. In: *Proceedings of the Institution of Electrical Engineers* 119 (1972), Nr. 12, S. 1733–1739
- [45] GODFREY, K. R.; MOORE, D. J.: Identification of processes having direction-dependent responses, with gas-turbine engine applications. In: *Automatica* 10 (1974), Nr. 5, S. 469–478
- [46] GODFREY, K. R.; TAN, A. H.; BARKER, H. A.; CHONG, B.: A survey of readily accessible perturbation signals for system identification in the frequency domain. In: *Control Engineering Practice* 13 (2005), November, Nr. 11, S. 1391–1402
- [47] GODWARD, T.; SCHILLING, H.; SCHAUM, S.: Use of Dynamic Testing and Modeling in the Engine Calibration Process. In: RÖPKE, K. (Hrsg.): *Design of Experiments (DoE) in Engine Development VII*, Expert-Verlag, Renningen, 2013, S. 211–230
- [48] GOLUB, G. H.; SAUNDERS, M. A.: Linear least squares and quadratic programming / Stanford University. 1969. – Technical Report CS 134. – Computer Science Department
- [49] GOLUB, G. H.; VAN LOAN, C. F.: An Analysis of the Total Least Squares Problem. In: *SIAM Journal on Numerical Analysis* 17 (1980), Nr. 6, S. 883–893
- [50] GSCHWEITL, K.; PFLUEGL, H.; FORTUNA, T.; LEITHGOEB, R.: Increasing the efficiency of model-based engine applications through the use of Cameo online DoE toolbox. In: *ATZ worldwide eMagazines* 17–20 (2001), Nr. 7-8, S. 274
- [51] GUNDLACH, C.: *Entwicklung eines ganzheitlichen Vorgehensmodells zur problemorientierten Anwendung der statischen Versuchsplanung*, Universität Kassel, Dissertation, 2004
- [52] GUTMAN, P.-O.: Stabilizing controllers for bilinear systems. In: *IEEE Transactions on Automatic Control* 26 (1981), Nr. 4, S. 917–922
- [53] HAFNER, M.: *Modellbasierte stationäre und dynamische Optimierung von Verbrennungsmotoren am Motorenprüfstand unter Verwendung neuronaler Netze*. Fortschritt-Berichte VDI: Nr. 482 in Reihe 12, VDI, Düsseldorf, 2001

- [54] HAFNER, M.: Effiziente Applikation der Motorsteuerung mit dynamischen Modellen (Efficient Calculation of Engine Control Maps by Means of Dynamic Models). In: *at - Automatisierungstechnik* 51 (2003), Nr. 5/2003, S. 213–220
- [55] HAFNER, M.; SCHÜLER, M.; NELLES, O.; ISERMANN, R.: Fast Neural Networks for Diesel Engine Control Design. In: *Control Engineering Practice* 8 (2000), Nr. 11, S. 1211–1221
- [56] HALTON, J. H.: On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. In: *Numerische Mathematik* 2 (1960), Nr. 1, S. 84–90
- [57] HAMETNER, C.; JAKUBEK, S.: Local model network identification for online engine modelling. In: *Information Sciences* 220 (2013), S. 210–225
- [58] HARTMANN, B.; EBERT, T.; NELLES, O.: Model-Based Design of Experiments based on Local Model Networks for Nonlinear Processes with Low Noise Levels. In: *American Control Conference*. San Francisco: American Automatic Control Council, Juli 2011, S. 5306–5310
- [59] HAUKAP, C.: *Steuerung und Regelung von KFZ-Antriebssträngen*.
https://www.mdt.tu-berlin.de/fileadmin/fg184/Lehre/SteuRegKFZ_Antriebe/v19.pdf, Abruf: 31. Juli. 2016. – 9. Vorlesung: Motorprüfstände und Prüfstandsautomatisierung
- [60] HEYWOOD, J. B.: *Internal combustion engine fundamentals*. McGraw-Hill New York, 1988
- [61] HOERL, A. E.; KENNARD, R. W.: Ridge Regression: Biased Estimation for Nonorthogonal Problems. In: *Technometrics* 12 (1970), Nr. 1, S. 55–67. – ISSN 00401706
- [62] ISERMANN, R.: *Identifikation dynamischer Systeme I*. Springer, Berlin, 1988
- [63] ISERMANN, R.: *Identifikation dynamischer Systeme II*. Springer, 1988
- [64] ISERMANN, R.: *Digital control systems: vol. 1: fundamentals, deterministic control*. Springer, Berlin, 1989
- [65] ISERMANN, R.: *Regelungstechnik I*. Shaker, 2002
- [66] ISERMANN, R.: *Modellgestützte Steuerung, Regelung und Diagnose von Verbrennungsmotoren*. Springer, Berlin, 2003
- [67] ISERMANN, R.: *Mechatronische Systeme: Grundlagen*. Springer, Berlin, 2007
- [68] ISERMANN, R.: *Engine Modeling and Control*. Springer, Berlin, 2014

- [69] ISERMANN, R.; KÖTTER, H.; SCHREIBER, A.: Vermessungsstrategien zur Modellbildung von Verbrennungsmotoren mit Prüfständen. In: *AUTOREG - Steuerung und Regelung von Fahrzeugen und Motoren*. Baden-Baden: VDI, Düsseldorf, Februar 2008, S. 531–544
- [70] ISERMANN, R.; MÜNCHHOF, M.: *Identification of Dynamic Systems*. Springer, 2011
- [71] ISERMANN, R.; SCHREIBER, A.: Identification methods for experimental modeling of nonlinear combustion processes. In: *4th International Energy Conversion Engineering Conference and Exhibit*. San Diego, Californien: AIAA, Juni 2006, S. 1468–1485
- [72] JACKSON, J. E.: *A user's guide to principal components*. John Wiley & Sons, Hoboken, 2003
- [73] JAKUBEK, S.; HAMETNER, C.: Identification of neurofuzzy models using GTLS parameter estimation. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 39 (2009), Nr. 5, S. 1121–1133
- [74] JAKUBEK, S.; KEUTH, N.: A local neuro-fuzzy network for high-dimensional models and optimization. In: *Engineering Applications of Artificial Intelligence* 19 (2006), Nr. 6, S. 705–717
- [75] JOHNSON, M. E.; NACHTSHEIM, C. J.: Some Guidelines for Constructing Exact D-Optimal Designs on Convex Design Spaces. In: *Technometrics* 25 (1983), Nr. 3, S. 271–277
- [76] JOLLIFFE, I. T.: *Principal Component Analysis*. Springer, 2002
- [77] JUNGE, T. F.: *On-line-Identifikation und lernende Regelung nichtlinearer Regelstrecken mittels neuronaler Netze*, Ruhr-Universität Bochum, Diss., 1999
- [78] KAMPELMÜHLER, F. T.; PAULITSCH, R.; GSCHWEITL, K.: Automatic ECU-Calibration - An Alternative to Conventional Methods / Society of Automotive Engineers. Society of Automotive Engineers, 1993 (930395). – SAE Technical Paper. – International Congress & Exposition
- [79] KENNEDY, J.; EBERHART, R.: Particle swarm optimization. In: *1995 IEEE International Conference on Neural Networks* Bd. 4 IEEE, 1995, S. 1942–1948
- [80] KIANIFAR, M.; CAMPEAN, L.; RICHARDSON, D.: Sequential DoE Framework for Steady State Model Based Calibration. In: *SAE International Journal of Engines* 6 (2013), Nr. 2, S. 570–583
- [81] KIEFER, J.; WOLFOWITZ, J.: Optimum Designs in Regression Problems. In: *The Annals of Mathematical Statistics* 30 (1959), Nr. 2, S. 271–294
- [82] KIRKPATRICK, S.; GELATT, C. D.; VECCHI, M. P.: Optimization by simulated annealing. In: *science* 220 (1983), Nr. 4598, S. 671–680

- [83] KLEIN, P.; LINSEN, R.: Moderne Methoden zur Motorgrundapplikation auf dem Prüfstand. In: *Elektronisches Management von Fahrzeugantrieben*. Wiesbaden: Haus der Technik, Oktober 2012, S. 1–14
- [84] KLÖPPER, F.: *Entwicklung und Einsatz modellgestützter Online-Methoden zur Parameteroptimierung von Verbrennungsmotoren am Prüfstand*, TU München, Dissertation, 2009
- [85] KNAAK, M.; SCHAUM, S.; RÖPKE, K.: Fast measurement by integrating model-based methods and advanced test-cell automation. In: *2006 IEEE International Conference on Control Applications*. München: IEEE, Oktober 2006, S. 2054–2059
- [86] KNÖDLER, K.: *Methoden der restringierten Online-Optimierung zur Basisapplikation moderner Verbrennungsmotoren*, Universität Tübingen, Dissertation, 2004
- [87] KNÖDLER, K.; POLAND, J.; FLEISCHHAUER, T.; MITTERER, A.; ULLMANN, S.; ZELL, A.: Modellbasierte Online-Optimierung moderner Verbrennungsmotoren - Teil 2: Grenzen des fahrbaren Suchraums. In: *MTZ - Motortechnische Zeitschrift* 64 (2003), S. 520–526
- [88] KOHLHASE, M.: *Brennraumdruckbasiertes Motormanagement für Otto und Dieselmotoren zur Verbrauchs- und Emissionsreduktion*. Fortschritt-Berichte VDI: Nr. 743 in Reihe 12, VDI, Düsseldorf, 2011
- [89] KÖTTER, H.: *Kennfeldoptimierung: Rechnergestützte Kennfeldoptimierung*. Frankfurt: FVV Abschlussbericht Vorhaben 836, Heft 831-2007, 2007
- [90] KÖTTER, H.: *Innovative Motorvermessung: Schnelle Motorvermessung mit optimierten Testsignalen*. Frankfurt: FVV Abschlussbericht Vorhaben 921, Heft 853-2008, 2008
- [91] KÖTTER, H.; SEQUENZ, H.; ISERMANN, R.: Ermittlung der Güte experimentell gewonnener Verbrennungsmotor-Modellen. In: *2. Internationales Symposium für Entwicklungsmethodik*. Wiesbaden: AVL, 2007, S. 95–112
- [92] KOWALCZYK, M.: *Online-Methoden - Optimierte schnelle Motorvermessung mit Online-Methoden zur Bestimmung von statischen und dynamischen Modellen*. Frankfurt: FVV Abschlussbericht Vorhaben 1035, Heft 987-2013, 2013
- [93] KOWALCZYK, M.: *Online-Methoden II - Erweiterung der Online-Vermessung um aktiv lernende Optimierungsmethoden zur Emissions- und Verbrauchsreduktion*. Frankfurt: FVV Abschlussbericht Vorhaben 1129, Heft 1043-2014, 2014
- [94] KRUSE, T.; KURZ, S.; LANG, T.: Modern Statistical Modeling and Evolutionary Optimization Methods for the Broad Use in ECU Calibration. In: *6th IFAC Symposium Advances in Automotive Control*, IFAC, 2010, S. 739–743
- [95] KRYKOVA, I.: *Evaluating of path-dependent securities with low discrepancy methods*, Worcester Polytechnic Institute, Masterarbeit, 2003

- [96] KUDER, J.; KRUSE, T.: Parameteroptimierung an Ottomotoren mit Direkteinspritzung. In: *MTZ - Motortechnische Zeitschrift* 61 (2000), Nr. 6, S. 378–385
- [97] LEONHARDT, S.: *Modellgestützte Fehlererkennung mit Neuronalen Netzen - Überwachung von Radaufhängungen und Diesel-Einspritzanlagen*. Fortschritt-Berichte VDI: Nr. 295 in Reihe 12, VDI, Düsseldorf, 1996
- [98] LEVENBERG, K.: A method for the solution of certain non-linear problems in least square. In: *Quarterly of Applied Mathematics* 2 (1944), S. 64–168
- [99] LI, B. X.: *Quasi-Monte-Carlo-Verfahren zur Bewertung von Finanzderivaten*, Goethe Universität Frankfurt am Main, Bachelorarbeit, 2011
- [100] LJUNG, L.: *System Identification: Theory for the user*. Prentice-Hall, Inc., 1999
- [101] MACKAY, D.: *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003
- [102] MARKOVSKY, I.; VAN HUFFEL, S.: Overview of total least-squares methods. In: *Signal Processing* Bd. 87, 2007, S. 2283–2302
- [103] MCCORMACK, A. S.; GODFREY, K. R.; FLOWER, J. O.: Design of multilevel multiharmonic signals for system identification. In: *IEE Proceedings – Control Theory and Applications* 142 (1995), Mai, Nr. 3, S. 247–252
- [104] MCKAY, M. D.; CONOVER, W. J.; BECKMAN, R. J.: A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. In: *Technometrics* 21 (1979), Nr. 2, S. 239–245
- [105] MICHALEWICZ, Z.: *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, 1992
- [106] MILLER, A. J.: *Subset Selection in Regression*. CRC Press, 2002
- [107] MITCHELL, T. J.: An Algorithm for the Construction of D-Optimal Experimental Designs. In: *Technometrics* 16 (1974), Nr. 2, S. 203–210
- [108] MITTERER, A.: *Optimierung vielparametrischer Systeme in der Kfz-Antriebsentwicklung: statistische Versuchsplanung und künstliche neuronale Netze in der Steuergeräteauslegung zur Motorabstimmung*, TU München, Dissertation, 2000
- [109] MITTERER, A.; ZUBER-GOOS, F.: Modellgestützte Kennfeldoptimierung - Ein neuer Ansatz zur Steigerung der Effizienz in der Steuergeräteapplikation. In: *ATZ - Automobiltechnische Zeitschrift* 3 (2000), S. 188–196
- [110] MONTGOMERY, D. C.: *Design and Analysis of Experiments*. 8. John Wiley & Sons, Hoboken, 2012

- [111] MOODY, J. E.: The Effective Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems. In: *Advances in Neural Information Processing Systems 4* (1991), S. 847–854
- [112] MOSER, W.: *Vergleichende Untersuchung von Brennraumsignalen bei Ottomotoren im Hinblick auf Motorregelung*, Technischen Universität Wien, Dissertation, 1992
- [113] MROSEK, M.; SEQUENZ, H.; ISERMANN, R.: Control Oriented NO_x and Soot Models for Diesel Engines. In: *6th IFAC Symposium Advances in Automotive Control*, IFAC, 2010, S. 234–239
- [114] MÜLLER, N.: *Adaptive Motorregelung beim Ottomotor unter Verwendung von Brennraumdrucksensoren*. Fortschritt-Berichte VDI: Nr. 545 in Reihe 12, VDI, Düsseldorf, 2003
- [115] MURAKAMI, Y.; KUSUKI, H.; NAKAGAWA, T.; YAMAMOTO, H.; LEITHGÖB, R.; RAINER, A.; MARTINI, E.: Slow dynamic slopes (SDS) methodology applied to a stratified gasoline engine. In: RÖPKE, K. (Hrsg.): *Design of Experiments (DoE) in Engine Development IV*, Expert-Verlag, Renningen, 2007, S. 413–423
- [116] MYERS, R. H.: *Classical and Modern Regression with Applications*. Duxbury, 1990
- [117] NAUMANN, T.: *Wissensbasierte Optimierungsstrategien für elektronische Steuergeräte an Common-Rail-Dieselmotoren*, TU Berlin, Dissertation, 2002
- [118] NELDER, J. A.; MEAD, R.: A Simplex Method for Function Minimization. In: *Computer Journal* 7 (1965), Nr. 4, S. 308–313
- [119] NELLES, O.: LOLIMOT - Lokale, Lineare Modelle zur Identifikation nichtlinearer, dynamischer Systeme. In: *at - Automatisierungstechnik* 45 (1997), S. 163–174
- [120] NELLES, O.: *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer, 2001
- [121] NELLES, O.; FINK, A.; BABUSKA, R.; SETNES, M.: Comparison of two construction algorithms for Takagi-Sugeno fuzzy models. In: *International Journal of Applied Mathematics and Computer Science* 10 (2000), Nr. 4, S. 835–855
- [122] PIERBURG: *AGR-Ventil Opel/Vauxhall - Pierburg*.
http://www.pierburg-service.de/ximages/pg_si_1011_de_web.pdf, Abruf: 31. Juli. 2016
- [123] POLAND, J.: *Modellgestützte und Evolutionäre Optimierungsverfahren für die Motorentwicklung*, Eberhard-Karls-Universität Tübingen, Dissertation, 2002
- [124] PREPARATA, F. P.; SHAMOS, M. I.: *Computational Geometry. An Introduction*. Springer, Berlin, 1985

- [125] PRESS, W. H.; TEUKOLSKY, S. A.; VETTERLING, W. T.; FLANNERY, B. P.: *Numerical recipes: the art of scientific computing*. Cambridge University Press, 2007
- [126] RASMUSSEN, C. E.; WILLIAMS, C. K. I.: *Gaussian Processes for Machine Learning*. MIT Press, 2006
- [127] RENNINGER, P.: Schnelle Variationsraumbestimmung - Voraussetzung für eine effektive Versuchsplanung. In: *Motorenentwicklung auf dynamischen Prüfständen*. Wiesbaden: Haus der Technik, November 2004, S. 1–12
- [128] RENNINGER, P.; ALEKSANDROV, M.: Rapid Hull Determination: a new method to determine the design space for model based approaches. In: RÖPKE, K. (Hrsg.): *Design of Experiments (DoE) in Engine Development II*, Expert-Verlag, Renningen, 2005, S. 14–29
- [129] RITTER, H.; MARTINETZ, T.; SCHULTEN, K.: *Neuronale Netze*. Addison-Wesley, 1990
- [130] RÖPKE, K. (Hrsg.): *Design of Experiments (DoE) in der Motorenentwicklung: Praktische Anwendungen in der Automobil-und Zuliefererindustrie*. Expert-Verlag, Renningen, 2003
- [131] RÖPKE, K.; KNAAK, M.; NESSLER, A.; SCHAUM, S.: Rapid Measurement - Basic Combustion Engine Calibration In One Day? In: *MTZ worldwide Edition* 04 (2007), S. 16–19
- [132] ROSENQVIST, F.; TAN, A. H.; GODFREY, K.; KARLSTROM, A.: Direction-dependent system modeling approaches exemplified through an electronic nose system. In: *IEEE Transactions on Control Systems Technology* 14 (2006), Nr. 3, S. 526–531
- [133] SAMPSON, D.: Use of Model-Based Calibration in the Test Cell. In: RÖPKE, K. (Hrsg.): *Design of Experiments (DoE) in Engine Development V*, Expert-Verlag, Renningen, 2009, S. 23–30
- [134] SANTNER, T. J.; WILLIAMS, B. J.; NOTZ, W.: *The Design and Analysis of Computer Experiments*. Springer, Berlin, 2003
- [135] SCHLOSSER, A.; SCHÖNFELDER, C.; HENDRIKX, M.; PISCHINGER, S.; SENTIS, T.: Automated ECU-Calibration - Example: Torque Structure of Gasoline Engine. In: RÖPKE, K. (Hrsg.): *Design of Experiments (DoE) in Engine Development V*, Expert-Verlag, Renningen, 2009, S. 137–150
- [136] SCHMITT, M.: *Untersuchung zur Realisierung mehrdimensionaler lernfähiger Kennfelder in Großserien-Steuergeräten*. Fortschritt-Berichte VDI: Nr. 246 in Reihe 12, VDI, Düsseldorf, 1995
- [137] SCHREIBER, A.; ISERMANN, R.: Identification of Combustion Engines with Special Orthogonal Excitation Signals. In: *Dynamic Systems and Control Conference*, ASME, 2010, S. 673–678

- [138] SCHREIBER, A.; ISERMANN, R.; VOGELS, M.-S.: Dynamische Vermessung und Modellbildung hochdimensionaler Zusammenhänge im Verbrennungsmotor mit dynamischen Motorenprüfständen. In: 2. *Internationales Symposium für Entwicklungsmethodik*. Wiesbaden: AVL, 2007, S. 27–39
- [139] SCHROEDER, M.: Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. In: *IEEE Transactions on Information Theory* 16 (1970), S. 85–89
- [140] SCHÜLER, M.: *Stationäre Optimierung der Motorsteuerung von PKW-Dieselmotoren mit Abgasturbolader durch Einsatz schneller neuronaler Netze*. Fortschritt-Berichte VDI: Nr. 461 in Reihe 12, VDI, Düsseldorf, 2001
- [141] SCHÜLER, M.; HAFNER, M.; ISERMANN, R.: Einsatz schneller neuronaler Netze zur modellbasierten Optimierung von Verbrennungsmotoren. In: *MTZ - Motortechnische Zeitschrift* 61 (2000), Nr. 10, S. 704–711
- [142] SCHÜLER, M.; HAFNER, M.; ISERMANN, R.: Model-based optimization of IC engines by means of fast neural networks, Part 1: Modelling the engine- and emission behavior. In: *MTZ worldwide Edition* 61 (2000), Nr. 10/11, S. 10–11
- [143] SCHWARTE, A.: *Modellbasierte Fehlererkennung und Diagnose des Ansaug- und Abgassystems von Dieselmotoren*. Fortschritt-Berichte VDI: Nr. 634 in Reihe 12, VDI, Düsseldorf, 2007
- [144] SCHWARTE, A.; HACK, L.; ISERMANN, I.; NITZKE, H. G.; JESCHKE, J.; PIEWEK, J.: Automatisierte Applikation von Motorsteuergeräten mit kontinuierlicher Motorvermessung. In: *AUTOREG - Steuerung und Regelung von Fahrzeugen und Motoren*. Wiesloch: VDI, Düsseldorf, März 2004, S. 651–663
- [145] SCHWEFEL, H. P.: *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Birkhäuser Verlag, 1977
- [146] SEQUENZ, H.: *Emission Modelling and Model-Based Optimisation of the Engine Control*. Fortschritt-Berichte VDI: Nr. 1222 in Reihe 8, VDI, Düsseldorf, 2013
- [147] SEQUENZ, H.; KELLER, K.; ISERMANN, R.: Zur Identifikation mehrdimensionaler Kennfelder für Verbrennungsmotoren. In: *at - Automatisierungstechnik* 60 (2012), Nr. 6, S. 344–351
- [148] SEQUENZ, H.; MROSEK, M.; ISERMANN, R.: A Global-Local Emission-Model for NO_x and Soot Emissions of Turbocharged CR-Diesel Engines. In: *ASME Dynamic Systems and Control Conference*, 2010, S. 719–726
- [149] SOBOL, I. M.: On the distribution of points in a cube and the approximate evaluation of integrals. In: *USSR Computational Mathematics and Mathematical Physics* 7 (1967), Nr. 4, S. 86–112
- [150] SÖDERSTRÖM, T.; STOICA, P. G.: *System Identification*. Prentice Hall, 1989

- [151] STADLBAUER, M.; HAMETNER, C.; DEREGNAUCOURT, M.; JAKUBEK, S.; WINSEL, T.: Online Measuring Method Using an Evolving Model Based Test Design for Optimal Process Stimulation and Modelling. In: *2012 IEEE International Instrumentation and Measurement Technology Conference*. Graz: IEEE, 2012, S. 1314–1319
- [152] STOICA, P.; SELEN, Y.: Model-Order Selection: A review of information criterion rules. In: *IEEE Signal Processing Magazine* 21 (2004), Nr. 4, S. 36–47
- [153] SUNG, A.: *Dynamische Vermessungsmethoden in der Online-Optimierung moderner Verbrennungsmotoren*, Universität Tübingen, Dissertation, 2009
- [154] SUNG, A.; KLÖPPER, F.; MITTERER, A.; WACHTMEISTER, G.; ZELL, A.: Modellbasierte Online-Optimierung in der Simulation und am Motorenprüfstand. In: *MTZ - Motortechnische Zeitschrift* 68 (2007), Nr. 01, S. 42–48
- [155] TAN, A. H.: Direction-dependent systems - A survey. In: *Automatica* 45 (2009), Nr. 12, S. 2729–2743
- [156] TAN, A. H.; GODFREY, K. R.: Modeling of direction-dependent processes using Wiener models and neural networks with nonlinear output error structure. In: *IEEE Transactions on Instrumentation and Measurement* 53 (2004), Nr. 3, S. 744–753
- [157] TEOMITZI, H. E.: *Entwurf und Vergleich prozessangepasster Anregungssignale zur dynamischen Vermessung*, TU Darmstadt, Bachelorarbeit, 2014
- [158] THEIL, H.: *Economic Forecasts and Policy*. North-Holland, Amsterdam, 1961
- [159] TIBSHIRANI, R.: Regression Shrinkage and Selection via the Lasso. In: *Journal of the Royal Statistical Society* 58 (1996), Nr. 1, S. 267–288
- [160] TIKHONOV, A. N.; ARSENIN, V. Y.: *Solutions of ill-posed problems*. Winston Washington, DC:, 1977
- [161] TÖPFER, S.: Approximation nichtlinearer Prozesse mit Hinging Hyperplane Baummodellen. In: *at - Automatisierungstechnik* 4 (2002), S. 147–154
- [162] TÖPFER, S.: *Hierarchische neuronale Modelle für die Identifikation nichtlinearer Systeme*. Fortschritt-Berichte VDI: Nr. 705 in Reihe 10, VDI, Düsseldorf, 2002
- [163] TÖPFER, S.; NELLES, O.: Polynommodelle, Kennfelder und neuronale Netze. In: ISERMANN, R. (Hrsg.): *Modellgestützte Steuerung, Regelung und Diagnose von Verbrennungsmotoren*. Springer, Berlin, 2003, Kapitel 7, S. 103–119
- [164] TURNER, P.; MONTAGUE, G.; MORRIS, J.: Neural networks in dynamic process state estimation and non-linear predictive control. In: *Proceedings of the 4th IEE conference on artificial neural networks*, 1995, S. 284–289
- [165] ULLRICH, T.: *Untersuchungen zur effizienten interpolierenden Speicherung von nichtlinearen Prozeßmodellen und Vorsteuerstrategien*. Shaker-Verlag, Aachen, 1999

- [166] VOGT, M.; MÜLLER, N.; ISERMANN, R.: On-line adaptation of grid-based look-up tables using a fast linear regression technique. In: *Journal of Dynamic Systems, Measurement and Control* 126 (2004), März, Nr. 4, S. 732–739
- [167] WARD, M. C.; BRACE, C. J.; VAUGHAN, N. D.; CEEN, R.; HALE, T.; KENNEDY, G.: Investigation of SSweep"Mapping Approach on Engine Testbed / Society of Automotive Engineers. 2002 (2002-01-0615). – SAE Technical Paper. – 1–13 S. – SAE World Congress & Exhibition
- [168] WILLIMOWSKI, M.: *Verbrennungsdiagnose von Ottomotoren mittels Abgasdruck und Ionenstrom*. Shaker-Verlag, Aachen, 2003
- [169] WOLD, H.: Partial Least Squares. In: *Encyclopedia of Statistical Sciences* 6 (1985), S. 581–591
- [170] YAP, T. T. V.; TAN, A. H.; FOO, M. F. L.: Comparison of bilinear systems with direction-dependent systems. In: *UKACC Control Conference*, 2004, S. ID–031
- [171] YOSHIDA, S.; EHARA, M.; WATANABE, S.; KURODA, Y.: Rapid Boundary Detection for Model Based Diesel Engine Calibration. In: RÖPKE, K. (Hrsg.): *Design of Experiments (DoE) in Engine Development VI*, Expert-Verlag, Renningen, 2011, S. 36–55
- [172] ZIMMERSCHIED, R.: *Entwurf von Anregungssignalen für die Identifikation nichtlinearer dynamischer Prozesse*, TU Darmstadt, Diplomarbeit, 2002
- [173] ZIMMERSCHIED, R.: *Identifikation nichtlinearer Prozesse mit dynamischen lokal-affinen Modellen*. Fortschritt-Berichte VDI: Nr. 1150 in Reihe 8, VDI, Düsseldorf, 2008
- [174] ZIMMERSCHIED, R.; ISERMANN, R.: Regularisierungsverfahren für die Identifikation mittels lokal-affiner Modelle. In: *at - Automatisierungstechnik* 56 (2008), S. 339–349
- [175] ZIMMERSCHIED, R.; WEBER, M.; ISERMANN, R.: Static and Dynamic Measurements of Combustion Engines for Optimization of Control Mappings - A Brief Survey. In: *at - Automatisierungstechnik* 53 (2005), Nr. 2/2005, S. 87–94

Eigene Veröffentlichungen

- [1] ISERMANN, R.; KOWALCZYK, M.: Online Measurement And Modeling Of Combustion Engines At Test Benches. In: *23rd Aachen Colloquium Automobile and Engine Technology*, 2014
- [2] ISERMANN, R.; KOWALCZYK, M.; KUNKEL, F.; SCHREIBER, A.; SEQUENZ, H.: Possibilities for target-oriented application of dynamic models in engine calibration. In: *4. Internationales Symposium für Entwicklungsmethodik*. Wiesbaden: AVL, 2011
- [3] KOWALCZYK, M.: *Online-Methoden - Optimierte schnelle Motorvermessung mit Online-Methoden zur Bestimmung von statischen und dynamischen Modellen*. Frankfurt: FVV Abschlussbericht Vorhaben 1035, Heft 987-2013, 2013
- [4] KOWALCZYK, M.: *Online-Methoden II - Erweiterung der Online-Vermessung um aktiv lernende Optimierungsmethoden zur Emissions- und Verbrauchsreduktion*. Frankfurt: FVV Abschlussbericht Vorhaben 1129, Heft 1043-2014, 2014
- [5] KOWALCZYK, M.; ISERMANN, R.: Online-Methods for the integration of Measurements, Modeling and Map Optimization at Engine Test Benches. In: *Design of Experiments (DoE) in Engine Development VII*, Expert-Verlag, Renningen, 2013, S. 107–124
- [6] KOWALCZYK, M.; ISERMANN, R.: Zielgerichtete Online-Vermessung zur iterativen Identifikation von stationären und dynamischen Verbrennungsmotormodellen. In: *AUTOREG - Steuerung und Regelung von Fahrzeugen und Motoren*, 2013
- [7] KOWALCZYK, M.; KUNKEL, F.; ISERMANN, R.: Online iterative real-time measurements and modeling of combustion engines. In: *13th Stuttgart International Symposium*, 2013, S. 515–533
- [8] KOWALCZYK, M.; YI, B.; ISERMANN, R.: Online Identification of Multivariable Nonlinear Processes with Iterative Structure Learning and Application to a Diesel Engine. In: *19th IFAC World Congress*, IFAC, 2014, S. 9216–9222
- [9] SCHREIBER, A.; KOWALCZYK, M.; ISERMANN, R.: Method for Dynamic Online Identification with Integrated Determination of Operating Boundaries. In: *Design of Experiments (DoE) in Engine Development VI*, Expert-Verlag, Renningen, 2011, S. 120–135

Betreute studentische Abschlussarbeiten

- [1] ECKART, A.: *Adaptive selbststellende Positionsregelung von Aktoren*, TU Darmstadt, Bachelorarbeit, 2013
- [2] HÜGELMANN, N.: *Untersuchung von Verfahren zur richtungsabhängigen Modellbildung von stationären und dynamischen Modellen*, TU Darmstadt, Bachelorarbeit, 2014
- [3] KLAPPAUF, L.: *Entwicklung, Untersuchung und Implementierung von Adaptionsstrategien innerhalb einer durchgehenden Online-Methodik zur Modellbildung an dynamischen Verbrennungsmotorprüfständen*, TU Darmstadt, Diplomarbeit, 2012
- [4] MEGEJ, N.: *Entwicklung einer onlinefähigen Rohdatenanalyse für dynamische Verbrennungsmotorprüfstände*, TU Darmstadt, Studienarbeit, 2011
- [5] PILLAS, J.: *Entwicklung einer modellbasierten Applikationsmethode für PKW-Antriebsstrangregelungen*, TU Darmstadt, Diplomarbeit, 2011
- [6] TEOMITZI, H. E.: *Entwurf und Vergleich prozessangepasster Anregungssignale zur dynamischen Vermessung*, TU Darmstadt, Bachelorarbeit, 2014
- [7] WACHENFELD, W.: *Modellgestützte Fehlerdiagnose eines Permanentmagnetenerregten Synchronmaschinen-Antriebs*, TU Darmstadt, Diplomarbeit, 2012
- [8] WANG, F.: *Iterative Identifikation semiphysikalischer globaler NOx-Emissionsmodelle von Dieselmotoren*, TU Darmstadt, Diplomarbeit, 2013
- [9] YI, B.: *Untersuchung von Verfahren zur rekursiven Modellbildung von stationären und dynamischen Verbrennungsmotormodellen*, TU Darmstadt, Masterarbeit, 2013

Online-Buchshop für Ingenieure

■ ■ VDI nachrichten

BUCHSHOP

Online-Shops



**Fachliteratur und mehr -
jetzt bequem online recher-
chieren & bestellen unter:
www.vdi-nachrichten.com/
Der-Shop-im-Ueberblick**



**Täglich aktualisiert:
Neuerscheinungen
VDI-Schriftenreihen**



Im Buchshop von vdi-nachrichten.com finden Ingenieure und Techniker ein speziell auf sie zugeschnittenes, umfassendes Literaturangebot.

Mit der komfortablen Schnellsuche werden Sie in den VDI-Schriftenreihen und im Verzeichnis lieferbarer Bücher unter 1.000.000 Titeln garantiert fündig.

Im Buchshop stehen für Sie bereit:

VDI-Berichte und die Reihe **Kunststofftechnik**:

Berichte nationaler und internationaler technischer Fachtagungen der VDI-Fachgliederungen

Fortschritt-Berichte VDI:

Dissertationen, Habilitationen und Forschungsberichte aus sämtlichen ingenieurwissenschaftlichen Fachrichtungen

Newsletter „Neuerscheinungen“:

Kostenfreie Infos zu aktuellen Titeln der VDI-Schriftenreihen bequem per E-Mail

Autoren-Service:

Umfassende Betreuung bei der Veröffentlichung Ihrer Arbeit in der Reihe Fortschritt-Berichte VDI

Buch- und Medien-Service:

Beschaffung aller am Markt verfügbaren Zeitschriften, Zeitungen, Fortsetzungsreihen, Handbücher, Technische Regelwerke, elektronische Medien und vieles mehr – einzeln oder im Abo und mit weltweitem Lieferservice

VDI nachrichten

BUCHSHOP

www.vdi-nachrichten.com/Der-Shop-im-Ueberblick

Die Reihen der Fortschritt-Berichte VDI:

- 1 Konstruktionstechnik/Maschinenelemente
 - 2 Fertigungstechnik
 - 3 Verfahrenstechnik
 - 4 Bauingenieurwesen
- 5 Grund- und Werkstoffe/Kunststoffe
 - 6 Energietechnik
 - 7 Strömungstechnik
- 8 Mess-, Steuerungs- und Regelungstechnik
 - 9 Elektronik/Mikro- und Nanotechnik
 - 10 Informatik/Kommunikation
 - 11 Schwingungstechnik
- 12 Verkehrstechnik/Fahrzeugtechnik
 - 13 Fördertechnik/Logistik
- 14 Landtechnik/Lebensmitteltechnik
 - 15 Umwelttechnik
 - 16 Technik und Wirtschaft
- 17 Biotechnik/Medizintechnik
- 18 Mechanik/Bruchmechanik
- 19 Wärmetechnik/Kältetechnik
- 20 Rechnerunterstützte Verfahren (CAD, CAM, CAE CAQ, CIM ...)
 - 21 Elektrotechnik
 - 22 Mensch-Maschine-Systeme
- 23 Technische Gebäudeausrüstung

ISBN 978-3-18-381012-3