

Modellbasierte Methoden zur transienten Optimierung von Antriebssystemen

Zur Erlangung des akademischen Grades eines
DOKTORS DER INGENIEURWISSENSCHAFTEN
(Dr.-Ing.)

von der KIT-Fakultät für Maschinenbau des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von

M.Sc. Niklas Ebert

Tag der mündlichen Prüfung 02.08.2023

Hauptreferent Prof. Dr. sc. techn. Thomas Koch
Korreferent Prof. Dr. Thomas H.W Bäck

Vorwort

Die Idee zu dieser Dissertation entstand in meiner Forschungstätigkeit als Doktorand im Team «Automatisierte Motoroptimierung» der Daimler AG in Stuttgart-Untertürkheim in Zusammenarbeit mit dem Institut für Kolbenmaschinen am Karlsruher Institut für Technologie. Die schriftliche Ausarbeitung ist dabei das Ergebnis verschiedener Entwicklungsprojekte, an denen ich mitgewirkt habe. Mein Dank gilt dabei einer Vielzahl von Kollegen, welche mir entsprechende Versuchsträger und Prüfstände zur Verfügung gestellt haben.

An erster Stelle gilt mein ganz besonderer Dank Herrn Prof. Dr. sc. techn. Thomas Koch für die Übernahme des Hauptreferats, sowie meiner fachlichen und persönlichen Betreuung am Karlsruher Institut für Technologie. Seine besondere Leidenschaft für die Weiterentwicklung von Antriebssystemen und die Reduzierung von Emissionen haben mich in besonderer Weise motiviert. Ebenso bedanke ich mich herzlich bei Herrn Prof. Dr. Thomas Bäck für die bereitwillige Übernahme des Koferats.

Ein vielfaches Dankeschön gebührt meinem Betreuer im Fachbereich Dr.-Ing. Frank Kirschbaum, dessen Unterstützung, fachliche Ratschläge sowie Fachkompetenz ich sehr geschätzt und genossen habe. Von der ursprünglichen Idee zu dieser Arbeit bis zur Verwirklichung war ich bei ihm immer in den allerbesten Händen.

Meinen Kollegen im Team «Automatisierte Motoroptimierung» danke ich für die ereignisreiche Zeit, ein angenehmes Arbeitsumfeld, tiefgehende Diskussionen sowie für ihre tatkräftige Unterstützung bei diesem Projekt. Hier gilt mein Dank besonders den Kollegen Jan-Christoph Goos und Sophia Leutenantsmayer bei der Unterstützung im mathematischen Bereich. Bei Philipp Werner und Dr.-Ing. Jens Vortisch bedanke ich mich für die engagierte Unterstützung bei der Prüfstandautomatisierung und bei Dr.-Ing. Richard Jakobi für die konstruktiven Anregungen und Ratschläge.

Ferner bedanke ich mich bei den Kollegen des KIT und der Daimler AG für die tolle Zusammenarbeit und Arbeitsatmosphäre.

Kurzfassung

Durch die Einführung des Konzepts der *Real driving emissions* (RDE) entstehen für die Motorenentwicklung neue Herausforderungen und damit auch für die etablierten Methoden der modellbasierten Kalibrierung. RDE-Ersatzzyklen weisen eine dynamische Charakteristik auf, die im Vergleich zum *Worldwide harmonized light duty vehicle test cycle* (WLTC) eine wesentliche Erhöhung der Emissionen nach sich zieht. Ist die Motorenkalibrierung des Betriebszustands mittels modellbasierter Kalibrierung in der Vergangenheit bevorzugt stationär erfolgt, steigt durch RDE die Bedeutung dynamischer Fahrmanöver. Eine stationäre Optimierung des Motorsteuergerätes reicht zur Einhaltung der Emissionsgrenzwerte nicht mehr aus. Ziel dieser Arbeit ist die Entwicklung einer modellbasierten Applikationsstrategie zur Optimierung von Antriebssystemen, um im dynamischen Motorbetrieb optimale Verbrauchs- und Emissionsgrößen zu erhalten.

Bei der Kalibrierung von Motorsteuergeräten werden die Sollwerte in Kennlinien und Kennfelder abhängig vom Betriebspunkt parametrisiert. Beispielsweise wird im Steuergerät der Kalibrierparameter „Einspritzzeitpunkt der Haupteinspritzung“ in einem Kennfeld definiert. Das Kennfeld ist über den beiden Führungsgrößen „Drehzahl“ und „Last“ aufgespannt. Anstelle der von diesen beiden Führungsgrößen abhängigen Steuergerätefunktion wird ein dreidimensionaler Kennraum für die modellbasierte Optimierung von transienten Emissionen eines Antriebssystems vorgeschlagen. Dabei wird die bestehende Definition des Betriebspunktes von einem zweidimensionalen Wertepaar auf ein dreidimensionales Wertetripel überführt. Unter Beibehaltung der bereits bestehenden Führungsgrößen wird als geeignete dritte Führungsgröße der Lastgradient identifiziert, der einen bedeutenden Einfluss auf die Höhe der Emissionen hat. Mittels statistischer Methoden erfolgen Voruntersuchungen, um den Versuchsaufwand zu reduzieren und die Wahl der Parameter für die Untersuchung vorab zu selektieren. Durch den dynamischen Charakter von RDE-Ersatzzyklen im Vergleich zu quasistationären Labortests steigen die nichtlinearen statistischen Zusammenhänge zwischen Motor- und Emissionssignalen. Da Emissionsmassenströme mittels Multiplikation der gemessenen Gaskonzentrationen mit dem Abgasmassenstrom berechnet werden, liefert die Arbeit zudem einen Beitrag zur Ermittlung von Gaslaufzeiten.

Abstract

The introduction of the Real driving emissions (RDE) presents new challenges in the area of model-based calibration for engine development. RDE cycles have a transient characteristic that results in a significant increase in emissions compared to the worldwide harmonized light duty vehicle test cycle (WLTC). While the engine calibration of the operating state by means of model-based calibration was preferably carried out stationary in the past, the significance of transient driving maneuvers increases due to RDE. A stationary optimization of the engine control unit is no longer sufficient to comply with the emission limit values. The aim of this work is the development of a model-based application strategy for the optimization of drive systems in order to obtain optimal consumption and emission values in dynamic engine operation.

When calibrating engine control units, the target values are parameterized in characteristic curves and maps depending on the operating point. For example, the calibration parameter injection time of the main injection is defined in a map in the control unit. The map value is depending on the two reference variables speed and load. Instead of the control unit function determined by the two reference variables speed and load, a three-dimensional characteristic space is proposed for the model-based optimization of transient emissions of a drive system. The existing definition of the operating point is transferred from a two-dimensional value pair to a three-dimensional value triple and examined on the test bench. The load gradient, which has a significant influence on the level of emissions, is identified as a suitable third reference variable. Using statistical methods, preliminary examinations are carried out in order to reduce the experimental effort and to select the input main variables for the examination. Due to the transient characteristics of RDE cycles compared to laboratory tests, the non-linear statistical correlation between engine and emission signals increase. Since emission signals are calculated by multiplying the measured gas concentrations and the exhaust gas mass flow, this work also makes a contribution to the treatment of gas transport delay.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Emissionen im praktischen Fahrbetrieb	1
1.2	Modellbasierte Kalibrierung	3
1.3	Wissenschaftliche Fragestellung	5
2	Stand der Technik der modellbasierten Kalibrierung	7
2.1	Empirische statische Modellbildung	7
2.1.1	Polynommodelle	8
2.1.2	Gauß-Prozess-Modelle	9
2.1.3	Lokale Modellnetze	10
2.2	Nichtlineare dynamische Modellierung	12
2.2.1	Modellkonfiguration in der Identifikation	13
2.2.2	Zeitdiskrete Neuronale Netze	16
2.2.3	Anregungssignale für dynamische Systeme	19
2.3	Analytische Kennfeldglattheit	20
2.3.1	Bilineare Interpolation von Betriebspunkten	21
2.3.2	Glattheitsmaß für Kennfeldstützstellen	23
2.3.3	Kennfeldglattheit unter Vorgabe exakter Punkte	24
3	Beiträge zur modellbasierten Kalibrierung	26
3.1	Methoden der Datenanalyse	26
3.1.1	Empirische Korrelationskoeffizienten	27
3.1.2	Verallgemeinerte Transinformation	28
3.1.3	Anwendungsbeispiele der Datenanalyse	31
3.2	Methoden der Sensitivitätsanalyse	39
3.2.1	Varianzbasierte Methoden	40
3.2.2	Visuelle Erklärbarkeitsansätze	42
3.2.3	Anwendungsbeispiele der Sensitivitätsanalyse	43
3.3	Methoden der Totzeitschätzung	49
3.3.1	Schätzung im Zeitbereich	49
3.3.2	Explizite Parameterschätzung	50
3.3.3	Verifikation und Vergleich der Schätzmethoden	53

3.4	Methodischer Ansatz zur RDE-Kalibrierung	57
3.4.1	Werkzeugkette der modellbasierten Kalibrierung	58
3.4.2	Aktueller Stand der Forschung zur Optimierung des transien- ten Betriebs	62
3.4.3	Erweiterung des Betriebspunkts um eine dynamische Füh- rungsgröße	63
3.4.4	Potenzialabschätzung der vorgeschlagenen Erweiterung	65
3.4.5	Bewertung der Bandbreite der Sensorik	68
3.5	Zusammenfassung der methodischen Untersuchung	71
4	Optimierung von Emissionen im transienten Fahrbetrieb	72
4.1	Auswahl der Kalibrierparameter und Versuchsplanung	72
4.1.1	Sensitivitätsanalyse der Kalibrierparameter	73
4.1.2	Auswahl der Lastgradienten	74
4.1.3	Versuchsplanung des Kalibrierparameters	77
4.2	Modellierung, Optimierung und Validierung	80
4.2.1	Statische Modellierung	82
4.2.2	Nichtlineare dynamische Modellierung	85
4.2.3	Optimierung der empirischen Modelle	86
4.2.4	Validierung und Interpretation der Ergebnisse	87
4.3	Erweiterung des Betriebspunktes	91
5	Zusammenfassung und Ausblick	92
	Literatur	95
	Nomenklatur	xi
	Abbildungsverzeichnis	xiv
	Tabellenverzeichnis	xv
A	Ergänzende Abbildungen der Optimierung des transienten Betriebs	xvi
B	Grundlagen der Stochastik	xxvi
B.1	Kenngößen empirischer Verteilungen	xxvi
B.2	Fehlermaße	xxvi
B.3	Rangtransformation	xxviii
B.4	Korrelationsintegral	xxx
B.5	Herleitung der Sensitivitätsanalyse nach Sobol	xxx
B.5.1	Monte-Carlo-Schätzung der varianzbasierten Sensitivitätsindizes	xxxii
B.5.2	Fourier Amplitude Sensitivity Test	xxxiii

C	Weiterführende Gleichungen zum Stand der Technik	xxxv
C.1	Statistische Versuchsplanung	xxxv
C.2	Anpassung der Hyperparameter des Gaußprozess Modells	xxxvii
C.3	Hierarchische Optimierung von lokalen Modellnetzen	xxxviii
C.4	Bestimmung von dynamischen Modellordnungen	xl
C.5	Numerische Optimierung	xlii
C.5.1	Partikelschwarmoptimierung	xliii
C.5.2	Multikriterielle Optimierung	xliv
C.5.3	Innere-Punkte Optimierung	xlvi
C.5.4	Formulierung der Kostenfunktion	l
C.6	Fehlerfunktion des Gütekriteriums der Kennfelderstellung	li
C.7	Systemidentifikation linearer dynamischer Modelle	lii
C.8	Modellgleichungen der Zufallssignale der Totzeitschätzung	liv

1 Einleitung

Die Motorenentwicklung wird derzeit von Diskussionen rund um den Klimawandel und die ersten Fahrverbote in verkehrsreichen Ballungsräumen bestimmt. Laut der Internationalen Energieagentur sind erkennbare Fortschritte in der Luftqualität nachweisbar, so sei beispielsweise in Deutschland über den Zeitraum von 1990 – 2019 der Kohlenstoffdioxidausstoß (CO₂-Ausstoß) um über 35 % verringert worden [123]. Ursachen hierfür sind v. a. wirtschaftliche Umstrukturierung sowie eine aktive Klimaschutzpolitik der Bundesregierung [123]. Der Trend der weltweiten CO₂-Emissionen verzeichnet dagegen ein Wachstum von über 35 % im Zeitraum von 1990 – 2015 [14]. Um die Ziele des Klimaschutzabkommens von Paris [118] zu erreichen, sind im Transportsektor weitere Maßnahmen erforderlich. Die Ergänzung von Verbrennungskraftmaschinen durch elektrische Antriebe und eine steigende Anzahl an Elektrofahrzeugen sind hier u. a. die Vorschläge, um die Emissionsziele bis 2025 zu erfüllen [129, S. 1].

1.1 Emissionen im praktischen Fahrbetrieb

Um die Umweltbelastungen durch die von Verbrennungsmotoren emittierten Schadstoffe zu verringern, werden die zulässigen Grenzwerte immer weiter verschärft [97, S. 170]. Speziell für Stickstoffoxid-Emissionen (NO_x) treten in Städten und Ballungsräumen an verkehrsnahen Messstellen im Tagesmittel Grenzwertüberschreitungen auf. Hier bescheinigt das Umweltbundesamt dem Straßenverkehr trotz eines fallenden Trends von NO_x-Emission einen bedeutenden Anteil [64]. Mit der Erweiterung der Euronorm *Euro 6c* [1] Mitte 2017 für Neutypen wird der im Jahr 1992 definierte synthetische Testzyklus der Abgaszertifizierung *New European driving cycle* (NEDC) durch den *Worldwide harmonized light duty vehicle test cycle* (WLTC) und *Real driving emissions* (RDE) ersetzt. Der WLTC kombiniert reale Fahrabschnitte zu einem standardisierten, laborgestützten Rollentest [3]. In der Vergangenheit sind die Optimierungen hinsichtlich der Schadstoffemissionen auf Grundlage der durch die Gesetzgebung festgeschriebenen synthetischen Fahrzyklen erfolgt. Mit der Einführung der RDE im September 2017 sind erstmalig in Europa Emissionslimits im realen Fahrbetrieb reglementiert [3]. Eine Übersicht über die verschiedenen Euronormen 6b, 6c, 6d_{TEMP} und 6d mit den jeweiligen zur Zertifizierung verwendeten Fahrzyklen gibt Tab. 1.1.

Tabelle 1.1 Zeitleiste der Einführung der Euronorm 6 und der zugehörigen Fahrzyklen sowie der Konformitätsfaktoren für Partikelanzahl (PN) und NO_x , ergänzt nach [80]

2014	2015	2016	2017	2018	2019	2020	2021	2022
	NEDC				WLTC und RDE			
	Euro 6b			Euro 6c		Euro 6d _{TEMP}		Euro 6d
					PN CF = 1,5			
		NO _x -Monitoring			NO _x CF = 2,1		CF = 1,5	

Die Zertifizierung im Rahmen der Euronorm 6b erfolgt durch den NEDC. In einer Übergangsphase, bezeichnet als Euronorm 6c, wird der Fahrzyklus durch den WLTC ersetzt. Seit der Einführung der temporären (TEMP) Norm Euro 6d_{TEMP} sind die Grenzwerte unter den vorläufigen Konformitätsfaktoren (CF) für die RDE gültig. Der CF entspricht einer maximal zulässigen realen Emissionsüberschreitung und ist für NO_x mit der Euronorm 6d seit 2019 gültig. Der CF bezieht Messtoleranzen u. a. gegenüber dem für RDE neu entwickelten portablen Emissionsmesssystem (PEMS) mit ein.

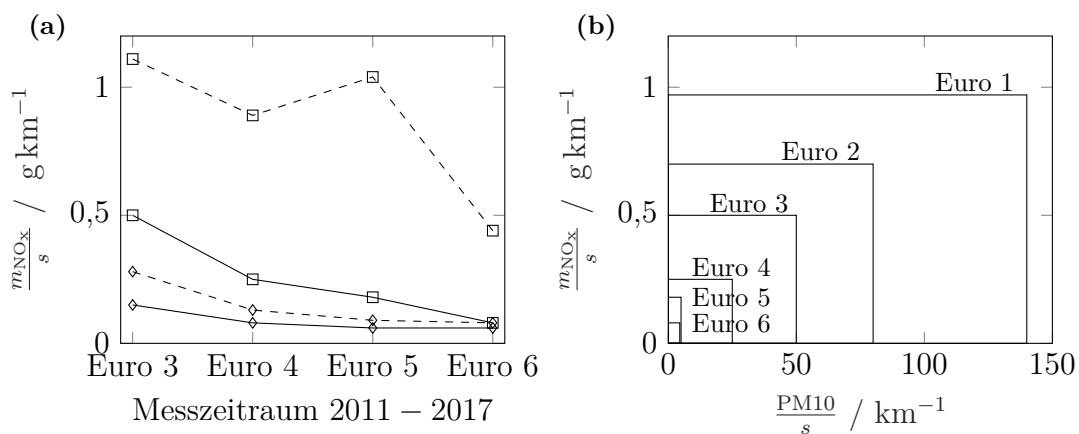


Abbildung 1.1 Diskrepanz zwischen Grenzwert (—) und realem Fahrverhalten (---) von PKW für Diesel- (□) und Ottomotoren (◇) [101, S. 2] (a). Entwicklung der Dieselgrenzwerte nach EU-Regularien von Euro 1 in 1992 bis zur Einführung von Euro 6 in 2014 für PKW (b) [69]

Mit der Ergänzung des Prüfverfahrens um RDE soll die Diskrepanz zwischen Labor- und Realemissionen von NO_x -Emissionen, dargestellt in Abb. 1.1a, reduziert werden [101, S. 2]. Hierbei sind die Abweichungen von Messungen im realen Fahrbetrieb und dem Grenzwert der jeweiligen Verschärfung der Emissionsgrenzwerte beim Ottomotor mit der Zeit marginal geworden. Mit der Euronorm 6 entsprechen die Werte der gemessenen Emissionen dem Level des Grenzwertes.

Für Dieselmotoren sinken die NO_x -Werte ebenfalls, die Abweichung zwischen Mess- und Grenzwerten bleiben allerdings erheblich. Die Grenzwerte der Euronormen von Euro 1 bis Euro 6 ist für NO_x - und PM_{10} -Partikel-Emissionen (Durchmesser bis $10\ \mu\text{m}$) von Dieselfahrzeugen in Abb. 1.1b dargestellt. Die PM_{10} -Grenzwerte zwischen Euro 1 und Euro 6 unterscheiden sich um über 96 %, die von NO_x , in Euro-norm 1 noch gemeinsam mit den Kohlenwasserstoffen (HC) reglementiert, sind um mehr als 91 % gesenkt [69].

Messstreuungen bei RDE-Straßentests sind nicht nur vom Fahrzeug, sondern auch von anderen Einflussgrößen abhängig. Solche sind beispielsweise der Geschwindigkeits- und Höhenverlauf, der Fahrer selbst, Messungenauigkeiten des Abgasmassenstrom für die Berechnung der Emission, die Umweltbedingungen sowie die PEMS-Anlage [3]. Die entsprechenden RDE-Gültigkeitsparameter können aus [3] entnommen werden. Eine zusätzliche Entwicklungsstufe von RDE kann aus den Empfehlungen von [101] für die Euro 7 Gesetzesstandards ab 2025 abgeleitet werden. Einige Kernpunkte für die Gesetzgebung und für die RDE lassen sich wie folgt zusammenfassen: Es wird eine weitere Verschärfung der Emissionsgrenzen am Beispiel Chinas und der USA sowie das Einbeziehen von CO , HC und Ammoniak-Emissionen bei PEMS-Fahrten empfohlen. Ferner existieren Bemühungen, um die Gültigkeitslimits wie Temperaturfenster der Abgasanlage, Länge des Stadtanteils und zulässige Höhenprofile von RDE-Fahrten mit PEMS zu erweitern. Zudem bestehen Anstrengungen, die Emissionslimits mit der Alterung der Abgasanlage über den Lebenszeitraum des Fahrzeuges zu gewährleisten [101].

1.2 Modellbasierte Kalibrierung

Das Betriebsverhalten von Antriebssystemen ist von einer Vielzahl von Steuergrößen abhängig [59, S. 7]. Die Steuergrößen werden abhängig von der Betriebsart oder -strategie unter Verwendung von verschiedenen Messgrößen in Steuergeräten in Echtzeit berechnet. Der Begriff der *Applikation* umfasst das Kalibrieren bzw. Parametrieren von Steuergeräteparametern [69, S. 3].

Der Paradigmenwechsel auf die RDE und zusätzlich die Prämissen der Gesetzeserweiterung Euro 7 erhöhen den Aufwand in der Entwicklung und in der Zertifizierung von Antriebssystemen mit elektronischer Motorsteuerung erheblich. Dies steht im Widerspruch zum Trend zu immer kürzer werdenden Entwicklungszyklen von modernen Verbrennungsmotoren. Getrieben durch neue Motor- und Triebstrangtechnologien steigt die Gesamtsumme der Parameter eines Motorsteuergerätes von ca. 10.000 im Jahr 2004 auf über 40.000 [65, 59, S. 8]. Durch die zunehmende Elektrifizierung des Antriebsstrangs ist mit einer weiteren Erhöhung der Anzahl an Kalibrierparametern, also der einzustellenden Betriebsparameter, zu rechnen.

Die Komplexität und Struktur der Software eines Motorsteuergerätes ist in Abb. 1.2 exemplarisch aufgezeigt. Die Darstellung der Software eines Diesel-Motorsteuergerätes unterscheidet Software von Daimler, von Zulieferern, Koppel-elemente und Funktionen mit Hardwarezugriff nach [71, S. 4].

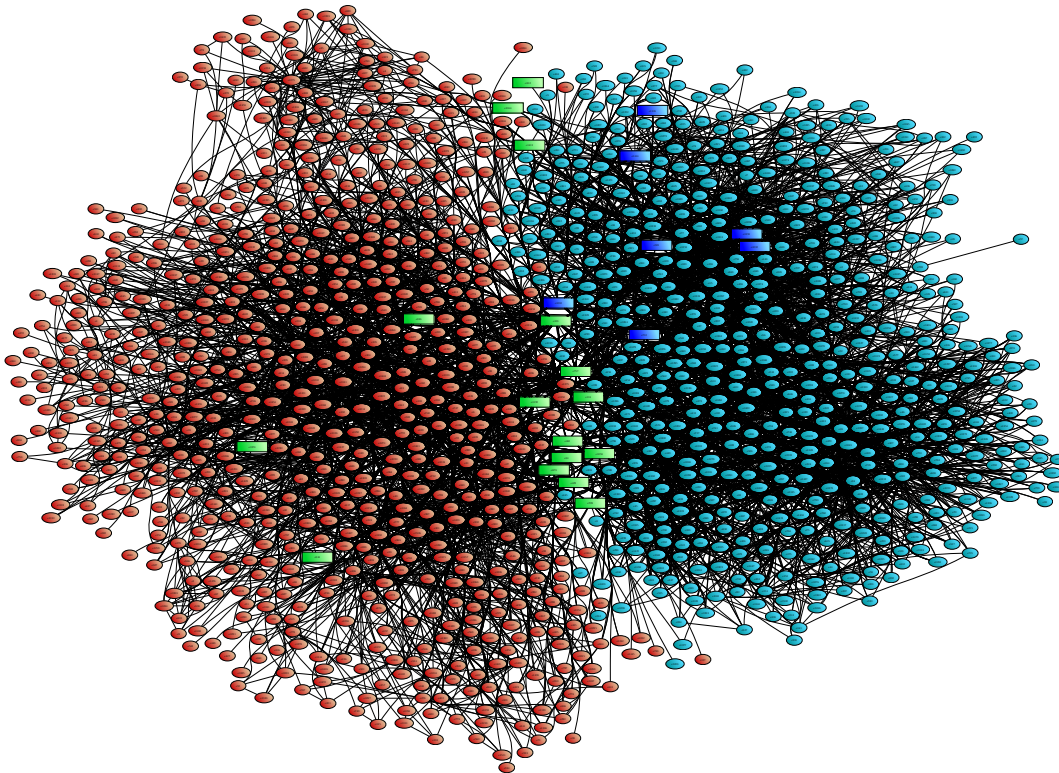


Abbildung 1.2 Struktur der Software des Motorsteuergeräts mit Software von Daimler (blau), Zulieferer-Software (rot) sowie Koppel- und hardwarenahen Funktionen (grüne und blaue Rechtecke) [71, S. 3]

Der Fokus der Applikation liegt auf der Optimierung der Kalibrierparameter des Steuergerätes. Diese umfasst den Verbrennungsmotor, die Abgasnachbehandlung und den in steigendem Maße elektrifizierten Antriebsstrang [3, S. 109]. In der Grundmotorenkalibrierung, d. h. der Optimierung von Rohemissionen und des Kraftstoffverbrauchs, haben sich hierbei modellbasierte Methoden etabliert [69]. Die modellbasierte Kalibrierung nutzt statistische Methoden, um die Versuchsumfänge zu reduzieren. Eine etablierte Werkzeugkette hat sich in der Vergangenheit als beste technische Praxis für Zyklusvorhersagen erwiesen. Diese Praxis ist besonders geeignet für transiente Zyklen wie beispielsweise den NEDC, da dieser häufige Anteile mit konstanter Fahrzeuggeschwindigkeit aufweist und in diesen Punkten der Motor stationär betrieben wird [69, S. 64]. Unter transient werden insbesondere zeitlich veränderliche Betriebszustände hinsichtlich Drehzahl und Last verstanden [69, S. 5].

1.3 Wissenschaftliche Fragestellung

Mit der Einführung der RDE-Ersatzzyklen ändert sich die Situation grundlegend, da kein eindeutig definierter Zyklus für die Kalibrierung vorliegt und die Annahme eines quasistationären Ersatztests bei der RDE nicht hinreichend erfüllt ist. Als Konsequenz führt die in der Vergangenheit bewährte Prozesskette nicht zu einer optimierten Kalibrierung von Verbrauch und Rohemission der in RDE-Fahrten auftretenden dynamischen Fahrzuständen. Die Positionspapiere [129, 128] empfehlen eine technologieoffene Weiterentwicklung von Antriebssystemen, in der die modellbasierte Kalibrierung einen wesentlichen Beitrag leisten kann. Ziel dieser Arbeit ist es, aufbauend auf bekannten Ansätzen der modellbasierten Kalibrierung und statistischen Auswertemethoden eine neue RDE-Auswertungs- und Kalibriermethodik zur Minimierung des Kraftstoffverbrauchs unter Einhaltung der gesetzlichen Emissionsgrenzen zu entwickeln. Dafür soll die bisherige Prozesskette in der Grundmotorenbedatung zur Zyklusvorhersage auf die Anforderungen von RDE angepasst werden. Die folgenden Fragestellungen definieren den Rahmen der wissenschaftlichen Arbeit:

- Im ersten Schritt soll der Begriff des Betriebspunkts von einer zweidimensionalen Drehzahl-Last-Softwarestruktur in eine um die Dynamik erweiterte Wertetripel-Struktur geprüft werden, um den Entwicklungsprozess von einem quasistationären Optimierungsziel auf ein transientes zu überführen. Kann eine **dynamische Größe** identifiziert werden und kann deren Einfluss auf die Zielgrößen der Kalibrierung quantifiziert werden? Lässt sich das Entwicklungspotenzial abschätzen und kann eine systematische Kalibriermethodik daraus abgeleitet werden?
- Ein Ziel des hier neu zu entwickelnden Ansatzes ist die Erweiterung der Führungsgrößen einer Steuererätedefunktion um eine dynamische Größe. Ist das **Dynamikverhalten** bzw. **Ansprechverhalten** der vorhandenen Emissionsmesseinrichtung in der Lage, den transienten Motorbetrieb zu untersuchen oder gibt es ab einer bestimmten Dynamik Grenzen der Messeinrichtungen?
- Der Versuchsaufwand der Grundmotorenkalibrierung steigt getrieben von neuen Softwarefunktionen zur Einhaltung der schärferen Grenzwerte kontinuierlich an. Eine einzelne Zielgröße in der Kalibrierung kann einen hochdimensionalen Eingangsparameterraum aufspannen. Gibt es Methoden der **Sensitivitätsanalyse**, mit denen der Eingangsraum auf die nur wesentlichen Kalibrierparameter reduziert werden kann? Lässt sich außerdem die Einflussgröße der Kalibrierparameter mit diesen Ansätzen feststellen?
- Zudem ist für die RDE die **Totzeitschätzung** (engl. *Time delay estimation*, TDE) für die Berechnung z. B. von Emissionsmassen aus gemessenen Gaskonzentrationen gesetzlich vorgeschrieben [2]. Die Totzeit ist die eine Zeitspanne zwischen einer Signaländerung am Systemeingang und der Signalantwort am Ausgang.

In stationären Messungen spielen Totzeiten in der Regel im System keine Rolle, jedoch sind sie in dynamischen Modellen von entscheidendem Einfluss auf die Modellgüte [120, S. 635]. Die exakte TDE zwischen Systemeingang und -ausgang stellt bei Signalen mit geringer statistischer Abhängigkeit eine besondere Herausforderung dar. Daher stellt sich die folgende Frage: Gibt es verlässliche Methoden aus dem Zeitbereich bzw. der Systemidentifikation für die exakte Schätzung der Totzeit zwischen den Messsignalen?

- Die Kalibrierung an einem Drehzahl-Last-Betriebspunkt erfolgt mittels stationärer Methoden. Im Transienten ändert sich der Betriebspunkt über der Zeit abhängig des Lastgradienten. Damit sind anstelle eines skalaren Wertes mehrere Kennfeldstützstellenwerte entlang der transienten Fahrspur in der Versuchsplanung zu berücksichtigen. Kann eine neue **Strategie in der Versuchsplanung** auch für transiente modellbasierte Kalibrieraufgaben entwickelt werden und ist diese Strategie auch in der Lage, die Dimension der Versuchsplanung auf einige wenige Freiheitsgrade zu reduzieren?
- Kann die modellbasierte **Optimierung von Kalibrierparametern im transienten Fahrbetrieb** weiterhin mit der klassischen, bewährten Methode einer statischen Regressionsmodellierung erfolgen? Dafür wird ein transienter Verlauf auf einen repräsentativen Skalar abgeleitet, der wiederum statisch modelliert wird. Können alternativ auch dynamische Modellierungsansätze aus dem Bereich des *Machine learning* in der modellbasierten Kalibrierung eine ausreichende Modellgüte erzielen?

Für die Optimierung von Emissionen im transienten Fahrbetrieb wird zunächst in Kap. 2 der Stand der Technik der modellbasierten Kalibrierung vorgestellt. Die Begriffe Modellierung und Kennfeldererstellung sowie Glättungsmaß werden eingeführt. Die Kennfeldglattheit ist für die Sicherheit von Stellgliedern bei Überanspruchung im dynamischen Betrieb sowie in Aspekten der Fahrbarkeit von besonderer Bedeutung. Darauf aufbauend werden in Kap. 3 Analysemethoden mit Anwendungsbeispielen vorgestellt. Die Verfahren sind für die Betrachtung der Momentendynamik notwendig. Die Sensitivitäts- bzw. Korrelationsanalyse quantifiziert statistische Zusammenhänge einzelner Parameter untereinander. Ferner ist für die Untersuchung von dynamischen Systemen mit Zeitverzögerung die Wahl einer geeigneten Methode der TDE ein Schwerpunkt des Kap. 3. Zudem ist die Totzeit notwendig bei der exakten Berechnung von Emissionsmassen in transienten Zyklen. Ziel der Kalibriermethodik ist die Reduktion der Emissionen in transienten Fahrmanövern. Daher wird das Potenzial einer Erweiterung der Führungsgrößen in der Motorsteuergerätefunktion um eine dynamische Größe evaluiert. Die maximale Grenzfrequenz, welcher die Sensorik noch folgen kann, wird hinsichtlich der transienten Kalibrierung schließlich untersucht. Im Anschluss daran wird in Kap. 4 die entwickelte Kalibrierstrategie in einer Messreihe überprüft. Abschließend folgt eine Empfehlung hinsichtlich einer zukünftigen, dreidimensional erweiterten Steuergerätefunktion abhängig von den Führungsgrößen Drehzahl, Last und Lastgradient.

2 Stand der Technik der modellbasierten Kalibrierung



Abbildung 2.1 Allgemeiner Prozess der modellbasierten Kalibrierung

In der Parametrierung von Motorsteuergeräten stellt die modellbasierte Kalibrierung ein Mittel dar, um die steigende Komplexität sowie den Zeitaufwand in der Kalibrierung mittels mathematischer Methoden zu beherrschen. Dabei sind von den Ingenieuren Kompromisse zwischen den Zielgrößen Verbrauch, Emissionen und Komfort bzw. Dynamik einzugehen. Die gefundenen Parameter werden in Form von skalaren Parametern, Kennlinien und Kennfeldern in den Steuergeräten hinterlegt [27, S. 1, 59, S. 12]. Ein Vorteil der modellbasierten Kalibrierung ist die Möglichkeit zur Optimierung der Parameter eines Modells anstelle eines Antriebssystems. Die zu applizierende bzw. zu parametrierende Steuergerätefunktion wird in Rückkopplung mit einem Streckenmodell simuliert. In einer Iterationsschleife erlaubt die modellbasierte Kalibrierung die Optimierung eines Gütekriteriums oder eines angestrebten Kompromisses [115, 8, 69, S. 7, 59, S. 13]. Der Standardvorgang der modellbasierten Kalibrierung ist durch die Schritte Versuchsplanung (engl. *Design of experiment*, DoE), Messung, Modellidentifikation, Optimierung, Validierung und Kalibrierung als linearer Prozess in Abb. 2.1 dargestellt. Der im Folgenden dargestellte Stand der Technik umfasst die grundlegenden Zusammenhänge und Eigenschaften der empirischen statischen Modellbildung, die nichtlineare dynamische Modellbildung sowie die Kennfelderstellung der Kalibrierung.

2.1 Empirische statische Modellbildung

Nach der experimentellen Untersuchung am Versuchsträger werden die empirischen Zusammenhänge zwischen den Eingangs- und den Ausgangsparametern in einem Modell abgebildet. In dieser Arbeit werden Polynommodelle, Gauß-Prozess-Modelle und lokale Modellnetze zur Beschreibung des Systemverhaltens vorgestellt. Die statische Modellbildung wird in dem DoE-Softwaretool *IAV EasyDoE* durchgeführt [55].

2.1.1 Polynommodelle

Ein Polynommodell ist ein beliebiger Modellierungsansatz in der modellbasierten Kalibrierung und aufgrund der Einfachheit der Anwendung weitverbreitet [85, 60, 61, 69, 8, 48]. Polynommodelle werden als Linearkombination zwischen dem Koeffizienten β und einem Regressor dargestellt. Der Regressor beschreibt einen Ausdruck, der von der Lage abhängt, an der das Modell ausgewertet werden soll. Ein Polynommodell ist ein sogenanntes lineares Modell, da es in den Koeffizienten linear ist. Der Begriff *linear* bezieht sich auf den Zusammenhang zwischen Koeffizienten- und Vorhersageraum, nicht den zwischen Eingangs- und Ausgangsgrößenraum [69, S. 26]. Ein n_p -dimensionales Polynommodell mit dem Eingangsgrößenvektor $\mathbf{u} = [u_1, u_2, \dots, u_{n_p}]^\top$ der Ordnung d ist definiert mit [85, S. 221 ff, 48, S. 9]

$$\hat{y}(\mathbf{u}) = \beta_0 + \sum_{i=1}^{n_p} \beta_i u_i + \sum_{i=1}^{n_p} \sum_{j=i}^{n_p} \beta_{i,j} u_i u_j + \sum_{i=1}^{n_p} \sum_{j=i}^{n_p} \sum_{k=j}^{n_p} \beta_{i,j,k} u_i u_j u_k + \dots \quad (2.1)$$

Die Anzahl der zu schätzenden Terme eines Polynommodells steigt mit zunehmender Eingangsdimension n_p und steigender Ordnung d aufgrund der Kreuzterme stark an. Die Modellierung wird daher zumeist auf niedrige quadratische oder kubische Terme beschränkt [115, S. 234]. Bei Überanpassung des Modells mit zu hohen Ordnungen neigen Polynommodelle zum Schwingen im Interpolationsbereich [48, S. 10]. Aus der Eigenschaft der Linearität erfolgt die Schätzung der Koeffizienten mit der Methode der kleinsten Quadrate (LS) [85, S. 222, 69, S. 26].

Das Schätzverfahren der Koeffizienten wird mit Strukturselektionsverfahren ergänzt, um komplexe Polynomstrukturen und eine Modellüberanpassung zu vermeiden. Variablen ohne erheblichen Einfluss auf die Modellgüte werden im Verfahren entfernt [85, S. 222]. Dabei identifizieren statistische Informationskriterien nicht signifikante Modellterme und deren Beitrag zum Modellvorhersagen. In dem DoE-Softwaretool *IAV EasyDoE* wird für die Regressorenauswahl der Algorithmus *Stepwise selection* verwendet. In jeder Iteration werden die selektierten Regressoren statistisch getestet, bevor ein neuer Regressor ausgewählt wird. Die als nicht signifikant deklarierten Modellterme werden aus dem Modell entfernt [85, S. 71 ff]. In Systembereichen mit starken Gradientenschwankungen besitzen Polynommodelle durch die in der Versuchsplanung vorgezogene Modellwahl bezüglich der Ordnung d eine beschränkte Flexibilität. Eine nachträgliche Erhöhung der Ordnung d ist nicht zulässig. An örtlich starken Gradientenänderungen kann eine Transformation der zu modellierenden Daten Abhilfe schaffen [69, S. 39]. Die *Box-Cox*-Transformation hilft beispielsweise, eine Schiefe der Verteilung der Messdaten zu reduzieren und die Modellgenauigkeit zu erhöhen [69, S. 39, 48, S. 147, 8, S. 67, 59, S. 65].

2.1.2 Gauß-Prozess-Modelle

Ein Gauß-Prozess (GP) ist eine Verallgemeinerung von ein- oder mehrdimensionalen Gaußverteilungen für unendlich viele Variablen [96, S. 2, 115, S. 290]. Die eindimensionale Gaußverteilung \mathcal{N} der Datenmenge \mathbf{U} ist durch den Mittelwert μ und die Standardabweichung σ vollständig definiert zu

$$\mathbf{U} \sim \mathcal{N}(\mu, \sigma^2) \quad \text{mit} \quad p(u) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left(-\frac{(u - \mu)^2}{2\sigma^2}\right). \quad (2.2)$$

Die Wahrscheinlichkeitsdichte für das Auftreten eines Punkts u in der Datenmenge \mathbf{U} beschreibt $p(u)$ [115, S. 290]. Eine n_p -dimensionale Gaußverteilung

$$\mathbf{U} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_u)$$

mit
$$p(\mathbf{u}) = \frac{1}{\sqrt{(2\pi)^{n_p} |\boldsymbol{\Sigma}_u|}} \exp\left(-\frac{1}{2}(\mathbf{u} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}_u^{-1}(\mathbf{u} - \boldsymbol{\mu})\right) \quad (2.3)$$

ist vollkommen definiert durch einen Mittelwertvektor $\boldsymbol{\mu}$ der Länge n_p , einer Kovarianzmatrix $\boldsymbol{\Sigma}_u$ der Dimension $n_p \times n_p$ und einer Determinante $|\boldsymbol{\Sigma}_u|$ [115, S. 291]. Eine Erweiterung für unendlich viele Variablen bildet einen mehrdimensionalen GP ab. Der mehrdimensionale GP Gl. 2.4 ist vollständig beschrieben durch die Substitution der endlichen Mittelwertfunktion $m(\mathbf{u})$ und der zugehörigen Kovarianzfunktion $k(\mathbf{u}, \mathbf{u}^*)$ in Gl. 2.3. Die Funktion $m(\mathbf{u})$ beschreibt den Erwartungswert einer beliebigen Funktion $f(\mathbf{u})$ und $k(\mathbf{u}, \mathbf{u}^*)$ die Kovarianz zweier Punkte im Faktorraum [115, S. 291]. Der GP wird durch

$$\hat{y}(\mathbf{u}) \sim \mathcal{GP}(m(\mathbf{u}), k(\mathbf{u}, \mathbf{u}^*)) \quad (2.4)$$

vollständig beschrieben [96, S. 13]. Eine Vorhersage \hat{y}^* für eine neue Eingangsmenge \mathbf{U}^* lässt sich durch die Funktionswerte $\hat{\mathbf{y}}$ an bekannten Punkten \mathbf{U} eindeutig bestimmen. Sind die Daten \mathbf{U} mit weißem Rauschen $\mathcal{N}(0, \sigma_n^2)$ überlagert, wird der Kovarianzausdruck $\mathbf{K}(\mathbf{U}, \mathbf{U})$ additiv um die Varianz des Rauschens $\sigma_n^2 \mathbf{I}$ erweitert. Die gemeinsame Verteilung ist definiert als [96, S. 16]

$$\begin{bmatrix} \hat{\mathbf{y}} \\ \hat{y}^* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}(\mathbf{U}, \mathbf{U}) + \sigma_n^2 \mathbf{I} & \mathbf{K}(\mathbf{U}, \mathbf{U}^*) \\ \mathbf{K}(\mathbf{U}^*, \mathbf{U}) & \mathbf{K}(\mathbf{U}^*, \mathbf{U}^*) \end{bmatrix}\right). \quad (2.5)$$

Hier bezeichnet $\mathbf{K}(\mathbf{U}, \mathbf{U}^*)$ die Kovarianzmatrix zwischen den Trainingsdaten \mathbf{U} und den Testdaten \mathbf{U}^* , gegeben sind also $\mathbf{K}(\mathbf{U}, \mathbf{U})$, $\mathbf{K}(\mathbf{U}^*, \mathbf{U})$ und $\mathbf{K}(\mathbf{U}^*, \mathbf{U}^*)$. Die Vorhersage \hat{y}^* unter den gegebenen Randbedingungen \mathbf{U} , $\hat{\mathbf{y}}$ und \mathbf{U}^* ermittelt die Verbundwahrscheinlichkeit als Normalverteilung mit einer Auswertung der Mittelwertfunktion und der Kovarianzmatrix mit folgenden Gl. [96, S. 16]:

$$\begin{aligned} \hat{y}^* | \mathbf{U}, \hat{\mathbf{y}}, \mathbf{U}^* &\sim \mathcal{N}(\hat{y}^*, \text{Cov}\{\hat{y}^*\}), \\ \hat{y}^* &= \mathbb{E}\{\hat{y}^* | \mathbf{U}, \hat{\mathbf{y}}, \mathbf{U}^*\} = \mathbf{K}(\mathbf{U}^*, \mathbf{U})[\mathbf{K}(\mathbf{U}, \mathbf{U}) + \sigma_n^2 \mathbf{I}]^{-1} \hat{\mathbf{y}}, \\ \text{Cov}\{\hat{y}^*\} &= \mathbf{K}(\mathbf{U}^*, \mathbf{U}^*) - \mathbf{K}(\mathbf{U}^*, \mathbf{U})[\mathbf{K}(\mathbf{U}, \mathbf{U}) + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{K}(\mathbf{U}, \mathbf{U}^*) \end{aligned} \quad (2.6)$$

Die Wahrscheinlichkeitsverteilung im Funktionsraum ist folglich von der A-priori-Bestimmung der Mittelwertfunktion $m(\mathbf{u})$ und der Kovarianzmatrix abhängig. Die Güte der Vorhersagegenauigkeit und Informationsdichte ist abhängig von der Anzahl an Datenpunkten [115, S. 295]. Zur Bestimmung der Kovarianzmatrix wird die quadratisch-exponentielle (engl. *Squared exponential* SE) Kernel-Funktion verwendet. Die Kernel-Funktion $k_{\text{SE}}(\mathbf{u}_p, \mathbf{u}_q)$ ist mit den drei Parametern der Signalvarianz σ_f^2 , der Längenskala ℓ_i in jede Eingangsdimension und der Varianz des Rauschens σ_n^2 definiert als [96, S. 19]

$$k_{\text{SE}}(\mathbf{u}_p, \mathbf{u}_q) = \sigma_f^2 \exp\left(-\sum_{i=1}^P \frac{(u_{p_i} - u_{q_i})^2}{2\ell_i^2}\right) + \sigma_n^2 \delta_{pq}. \quad (2.7)$$

Das Kronecker-Delta δ_{pq} entspricht dem Wert eins für den Fall $p = q$, andernfalls dem Wert null [96, S. 16]. Die Parameter $\sigma_f^2, \ell_i, \sigma_n^2$ der Kovarianzfunktion werden als Hyperparameter bezeichnet. Der Hyperparameter σ_f entspricht der Standardabweichung im Prozess und kontrolliert die vertikale Skalierung des Gauß-Prozess-Modells (GPM). Der Hyperparameter ℓ_i bestimmt die Längenskalierung jeder Eingangsdimension. Mit dieser Eigenschaft eignet sich die Längenskala ℓ_i folglich als Sensitivitätsmaß, denn der Hyperparameter ist für jede Eingangsdimension unterschiedlich groß [8, S. 32ff, 96, S. 125, 32]. Ist der Längenparameter ℓ_i klein im Verhältnis zu den übrigen Längenskalen, wird die Kernel-Funktion Gl. 2.7 sensitiv gegenüber der zugehörigen Eingangsvariable u_i . Dagegen ist die Kovarianzmatrix nahezu unabhängig von einer bestimmten Eingangsgröße u_i , wenn die Längenskala ℓ_i groß ist [96, S. 126]. Infolgedessen kann die Längenskala als Indikator für Modellinterpretationen genutzt werden. In einer Sensitivitätsanalyse bedeutet ein niedriger Wert der Längenskala ein hohes, ein hoher Wert dagegen ein geringes nichtlineares Verhalten [8, S. 32].

GPM sind flexibel in der Anwendung und bieten Vorteile in globalen experimentellen Anwendungen für das Vorhersagen von Nichtlinearitäten, Risiko- oder Sensitivitätsanalysen und Optimierungen [60, S. 181, 8, S. 30]. Außerdem weisen GPM durch die Kernel-Funktion Gl. 2.7 ein stabiles Verhalten bei Messdaten außerhalb der Modellvorhersage auf und streben zum Mittelwert [8, S. 41]. Ein GPM kann zudem um eine unendliche Anzahl an Basisfunktionen erweitert werden [8, S. 42, 115, S. 293].

2.1.3 Lokale Modellnetze

Lokale Modellnetze (LMN) sind exemplarisch in Abb. 2.2 dargestellt [nach 85, S. 342]. Als eine Klasse der Künstlichen Neuronalen Netze (KNN) sind LMN eine Erweiterung für Netze mit *Radial basis functions* (RBF). Jeder Netzknoten bildet dabei eine Realisierung eines lokalen linearen Modells ab. Dadurch sind LMN in der Lage, nichtlineare Funktionen stückweise mit mehreren linearen Modellen zu approximieren. LMN sind zur Identifikation nichtlinearer statischer und dynamischer Prozesse geeignet [85, S. 342, 48, S. 16, 121, S. 31].

Die jeweiligen lokalen Modelle können von beliebigen – linearen wie nichtlinearen – Funktionen beschrieben werden. Üblich ist die Wahl von linearen Ansätzen mit Polynommodellen. Dadurch können die Modellparameter mit linearen Methoden geschätzt werden. Der Ausgang \hat{y} eines LMN wird durch eine über die Gültigkeitsfunktion Φ_i gewichtete Summe aller einzelnen Ausgänge der LMN \hat{y}_i berechnet mit [85, S. 342]

$$\hat{y}(\mathbf{u}) = \sum_{i=1}^{n_M} \underbrace{(\beta_{i,0} + \beta_{i,1}u_1 + \beta_{i,2}u_2 + \dots + \beta_{i,n_p}u_{n_p})}_{\hat{y}_i(\mathbf{u})} \Phi_i(\mathbf{u}) \quad (2.8)$$

mit $\sum_{i=1}^{n_M} \Phi_i(\mathbf{u}) = 1.$

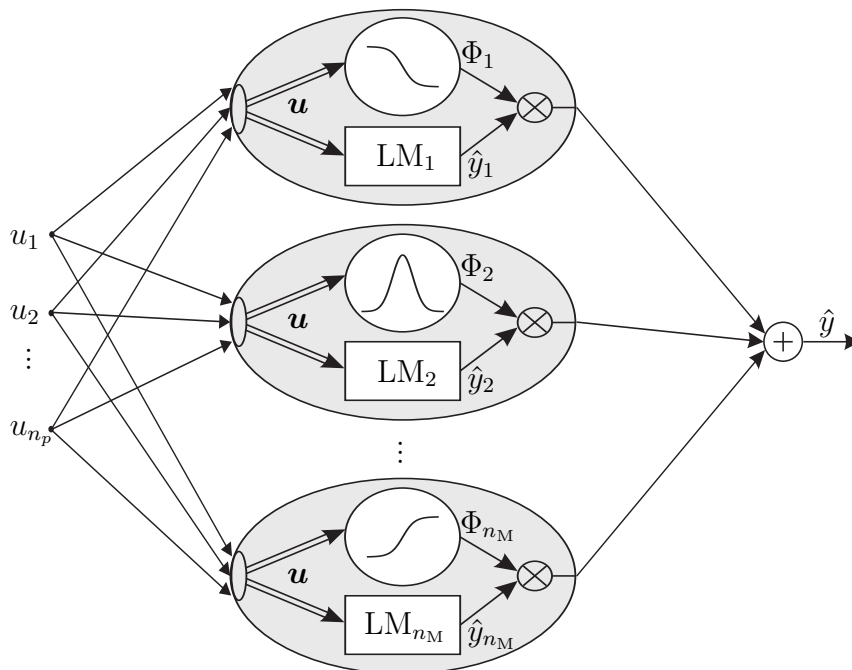


Abbildung 2.2 Lokales Modellnetz nach [nach 85, S. 342]. Das Aufsummieren der jeweils gewichteten lokalen Modelle ergibt den Modellausgang \hat{y} . Die Gültigkeitsfunktion oder Aktivierungsfunktion Φ_i gewichtet das zugehörige lokale Modell

Die Gültigkeitsfunktion oder Aktivierungsfunktion Φ_i beschreibt das Gebiet, in dem ein LMN gültig ist, normiert im Intervall $[0, 1] \in \mathbb{R}$. Die Summe aller n_M lokalen Modelle ergibt für jeden Eingangsvektor $\mathbf{u} = [u_1, u_2, \dots, u_{n_p}]^T$ mit n_p Eingängen genau eins [85, S. 342]. Diese Eigenschaft der Gültigkeitsfunktion Φ_i stellt die Interpretierbarkeit des Netzes sowie den Beitrag aller lokalen Modelle zu 100% sicher [48, S. 16]. Zwischen den einzelnen Modellnetzen wird stetig differenzierbar gewechselt und über die Gültigkeitsfunktion interpoliert [48, S. 14]. Typischerweise werden normierte Gauß- oder Sigmoidfunktionen verwendet. Sind die einzelnen Ausgänge \hat{y}_i konstant, entspricht die Modellgleichung der eines RBF-Netzes [48, S. 17].

Das Modell partitioniert sich stückweise, bewertet eine Struktur und passt die Nichtlinearität eines abzubildenden Prozesses an. In der Modellbildung muss ein Kompromiss zwischen einem stetigen Übergang der einzelnen Modellnetze und der Komplexität des Gesamtmodells gefunden werden. Je nach Wahl des Modellansatzes mit einem RBF-Netz oder einem Polynommodell erhöht sich in der Partitionierung des Modellraums die Anzahl der lokalen Modelle n_M (Neuronen) oder die Anzahl an Parametern pro lokalem Modell [48, S. 19].

Die Bestimmung der Parameter und die Partitionierung des LMN kann durch nichtlineare Optimierung erfolgen. In praktischer Anwendung sind heuristische Baumkonstruktions- und Clustering-Verfahren verbreitet [85, 48, 121, 30, 19]. In dem DoE-Softwaretool *IAV EasyDoE* erfolgt die Optimierung mit dem hierarchischen Optimierungsalgorithmus *Hierarchical local model tree* (HILOMOT) [48]. Dieser basiert auf dem *Local linear model tree* (LOLIMOT) Algorithmus und verwendet anstatt einer achsenorthogonalen Schnittoptimierung eine achsenschräge Partitionierung [48, S. 53]. Dadurch bietet HILOMOT einen flexibleren Ansatz als der von LOLIMOT und kann eine hochdimensionale Problemstellung auf die wirkliche Dimension der Nichtlinearität reduzieren [48, S. 74]. Die Gültigkeitsfunktion Φ_i ist nach dem HILOMOT-Algorithmus die Sigmoidfunktion. Diese weist im Vergleich zur Gaußfunktion weniger zu approximierende Parameter auf [48, S. 73].

2.2 Nichtlineare dynamische Modellierung

In der Regel werden statische Modelle als Systeme ohne Gedächtnis klassifiziert. Ein dynamisches Modell kann hingegen ein dynamisches Systemverhalten mathematisch abbilden. Im Unterschied zu statischen Modellen sind bei dynamischen die Ausgangsgrößen nicht nur abhängig vom Momentanwert der Eingangsgröße, sondern vom Verlauf des Systemeinganges oder von Informationen der Vergangenheit. Statische Modelle werden durch Funktionen beschrieben, dynamische Modelle durch Differenzgleichungen [72, S. 1]. Dynamische Modelle werden infolge dieser Eigenschaft als speicherfähige Systeme bezeichnet [12, S. 234]. Nichtlineare dynamische Modelle sind generell in folgender Form mit einer nichtlinearen Funktion f_{NL} für eine beliebige Menge an Eingängen n_m definierbar:

$$\hat{y}(k) = f_{NL}(\varphi(k)) = f_{NL}(\varphi_1, \varphi_2, \dots, \varphi_{n_m}) \quad (2.9)$$

Der Regressorvektor $\varphi(k)$ repräsentiert die Modelleingänge der nichtlinearen Funktion f_{NL} [85, S. 557]. Ferner lassen sich nichtlineare Ein- und Ausgangsmodellklassen in Modelle mit oder ohne Ausgangsrückkopplung aufteilen. Der Regressorvektor setzt sich aus vergangenen sowie ggf. aus derzeitigen Prozesseingängen $\varphi_i = u(k-i)$ und vergangenen $\varphi_i = y(k-i)$ Modellausgängen zusammen.

Folgende Modellstrukturen werden hier vorgestellt: das *nichtlineare Autoregressive with exogenous input-* (NARX)-Modell, das *nichtlineare Output error-* (NOE)-Modell und das *nichtlineare Finite impulse response-* (NFIR)-Modell. Die Modelle NARX und NOE besitzen eine Ausgangsrückkopplung, das NFIR-Modell nicht.

$$\text{NARX} : \boldsymbol{\varphi}(k) = [u(k-1), \dots, u(k-n_m), y(k-1), \dots, y(k-n_m)]^\top \quad (2.10)$$

$$\text{NOE} : \boldsymbol{\varphi}(k) = [u(k-1), \dots, u(k-n_m), \hat{y}(k-1), \dots, \hat{y}(k-n_m)]^\top \quad (2.11)$$

$$\text{NFIR} : \boldsymbol{\varphi}(k) = [u(k-1), u(k-2), \dots, u(k-n_m)]^\top \quad (2.12)$$

Die Unterscheidung der Ausgangsrückkopplungen von NARX- und NOE-Modellstruktur liegt in der Betrachtung des gemessenen Prozessausgangs $y(k-i)$ und des geschätzten Modellausgangs $\hat{y}(k-i)$. Bei der Systemidentifikation von linearen dynamischen Modellen liegt die Schwierigkeit in der Approximation der Parameterwerte. Bei der Schätzung von nichtlinearen dynamischen Modellen erweitert sich die Problemstellung auf die Wahl einer Funktion. Als nichtlineare Funktion f_{NL} ist jeder beliebige nichtlineare Modellansatz einsetzbar [85, S. 547].

2.2.1 Modellkonfiguration in der Identifikation

Neben dem Modellansatz muss zwischen unterschiedlichen Modellkonfigurationen dynamischer Modelle unterschieden werden. Die Parameter eines Modell lassen sich je nach gewählter Konfiguration als Simulator oder als Ein-Schritt-Prädiktor optimieren [85]. Das ist im Wesentlichen abhängig von den als Modelleingang gewählten Regressoren. Bei der Konfiguration als **Prädiktor** stehen neben den Systemeingängen $u(k)$ die bisherigen Messwerte $y(k)$ zur Verfügung. Bei der Konfiguration als **Simulator** ist das Modell ausschließlich von den bekannten Modelleingängen $u(k)$ abhängig [27, S. 14].

- Ein **Prädiktor** bestimmt die Ausgänge $y(k)$ in Abhängigkeit von bisher bekannten Messwerten. Die Dynamik wird in der Prädiktor-Konfiguration durch die vergangenen Beobachtungen außerhalb der Modellstruktur extern erzeugt. Das Modell ist statisch, da darin keine Zeitabhängigkeit abgebildet wird. Das Optimierungsziel des Verfahrens ist die Minimierung des Ein-Schritt-Prädiktionsfehlers des Modells. Eine Anforderung der Prädiktion ist, dass der Prozessausgang während des Vorgangs gemessen wird [27, S. 14].
- Im Gegensatz dazu ist ein **Simulator** ein freies Modell ohne Messdatenabgleich. Die rekurrente Struktur des Modells gestaltet sich erheblich aufwendiger in der Identifikation als bei der Prädiktion. In der Simulationskonfiguration wird bei der Optimierung der Modellparameter die Differenz zwischen Systemausgang und simuliertem Modellausgang minimiert. Die Simulation erfordert daher eine Sicherstellung der Modellstabilität [27, S. 14].

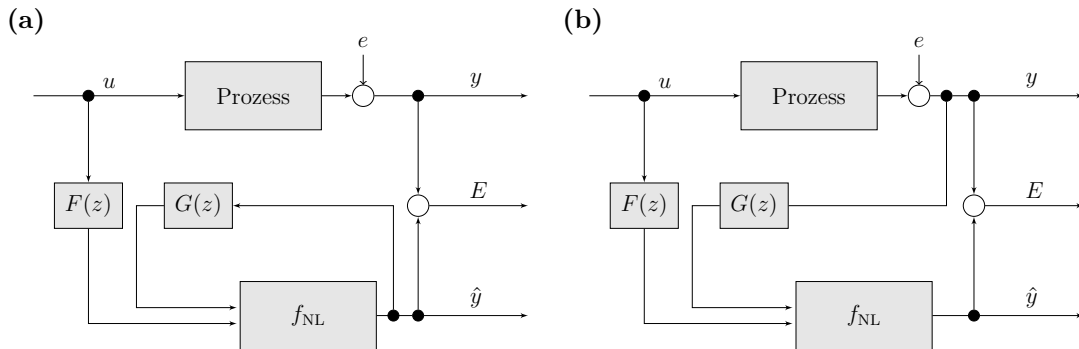


Abbildung 2.3 Nichtlineare Modellstrukturen des NOE-Modells in Simulator-Konfiguration (a) und des NARX-Modells in Prädiktor-Konfiguration (b). Eine NFIR Struktur ergibt sich bei Nichtberücksichtigung der Transferfunktion $G(z)$ [112, S. 246, 85, S. 556, 27, S. 54]

Die beiden Darstellungen der Simulator- und der Prädiktor-Konfigurationen sind in Abb. 2.3 mit den Transferfunktionen $F(z)$ und $G(z)$ dargestellt [112, S. 246, 85, S. 556, 27, S. 54]. Beim NOE-Modell werden die Ausgänge \hat{y} der nichtlinearen Funktion f_{NL} durch die Transferfunktion $G(z)$ verzögert auf die Netzeingänge zurückgeführt. Damit stellt die rekurrente Struktur einen Simulator dar (Abb. 2.3a). Für nichtlineare dynamische Modelle mit Ausgangsrückführung ergibt sich die in Abb. 2.3b dargestellte Struktur. Der gemessene Prozessausgang wird durch die Transferfunktion $G(z)$ verzögert auf den Eingang zurückgeführt. Diese Konfiguration entspricht dem NARX-Modell. Wenn der verzögerte Prozessausgang nicht zurückgeführt wird, ergibt sich das NFIR-Modell [112, S. 246].

Die praktische Anwendung einer Funktionsapproximation mit externer Dynamik, dem *Time delay neural network* (TDNN), ist der am meisten verbreitete nichtlineare dynamische Modell- und Identifikationsansatz [85]. Das Modell besteht aus zwei separaten Teilen: einem nichtlinearen, statischen Approximator und einer externen, dynamischen Filterbank. Die Darstellung mit externer Dynamik realisiert eine Differenzgleichung der Ordnung n_m und legt die Anzahl der Verzögerungsglieder fest [27, S. 15]. Als nichtlinearer Approximator wird in dieser Arbeit eine KNN-Topologie eingesetzt (siehe Kap. 2.2.2). Grundsätzlich werden die beiden Konfigurationen Prädiktion und Simulation zur Darstellung der Dynamik verwendet. Beide Konfigurationen werden durch die externe dynamische Filterbank realisiert [27, S. 15]. In Abb. 2.4 ist die TDNN Struktur eines NARX-Modells mit der Rückführung des gemessenen Ausgangs $y(k)$ als Eingangsgröße exemplarisch aufgezeigt [nach 85, S. 549].

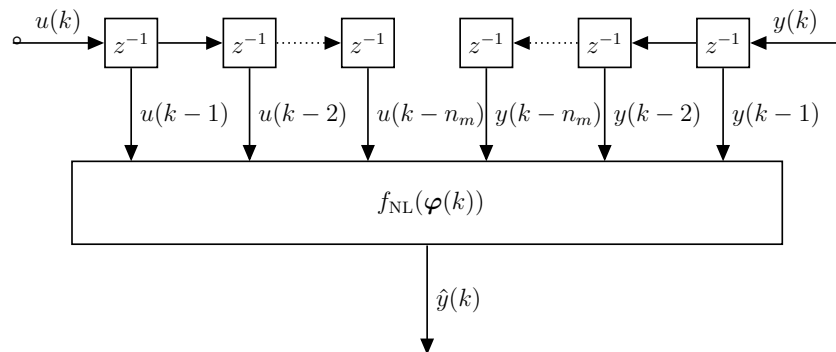


Abbildung 2.4 Darstellung eines nichtlinearen dynamischen Modells mit externer Dynamik (TDNN) einer NARX-Modellstruktur mit der Separierung in eine externe Filterbank mit Zeitverzögerungsgliedern und einen nichtlinearen statischen Approximator $f_{\text{NL}}(\varphi(k))$ [nach 85, S. 549]

Wesentlicher Vorteil der Nutzung des TDNNs ist eine moderate Anzahl an Modellparametern durch die Steuerung der Modellkomplexität über die festzulegenden Regressoren für die Filterbank. Zudem gibt es keine zu modellierende rekurrente Struktur [27, S. 16]. Aus der Definition der Ein-Schritt-Prädiktion bilden die Eingänge u und y der Filterbank eine NARX-Struktur ab wie in Abb. 2.4. Werden Modellausgänge \hat{y} verzögert in die Filterbank eingeführt, folgt die Parameterschätzung mit externer Dynamik in Simulationsanordnung. Dies entspricht der NOE-Modellstruktur. Folglich gestaltet sich die Parameterschätzung bei NARX-Modellen (Prädiktion) erheblich einfacher als bei NOE-Strukturen (Simulation), da das NOE-Modell rekurrente Bestandteile durch den verzögerten Modellausgang \hat{y} besitzt [85]. Bei NOE-Modellen wird der Simulationsfehler minimiert. Dies erfordert die Implementierung eines rekurrenten Lernalgorithmus'. NARX-Modelle sind bezüglich der Abtastung sensibel und betonen hohe Frequenzen verglichen mit NOE-Modellen stärker [27, S. 16]. Eine Erweiterung zu multivariablen Systemen gestaltet sich einfach, da beliebige Eingangsgrößen direkt in die Filterbank als Regressor eingeführt werden können [85].

NFIR-Modelle ohne eine Ausgangsrückkopplung haben eine hohe Anzahl an Regressoren. Die Wahl der Prozessordnung und die Berücksichtigung entscheidender Eingangsgrößen sind bestimmend für die Modellqualität. Die Bestimmung dynamischer Modellordnungen für Modelle ohne Ausgangsrückführung ist in Kap. C.4 im Anhang dargestellt [85, S. 574]. Der mit der Ordnung der Dynamik steigenden Größe des Regressorenvektors φ kann somit entgegengewirkt werden. Ein Nachteil von Modellen mit Ausgangsrückführung ist, dass die Ordnung n_m entscheidend für die Modellqualität ist. Es existiert bisher keine effiziente Methode zur Bestimmung der Ordnung n_m . Wenn die Ordnung für den Eingang n_a sich von der Ordnung des rückgekoppelten Ausgangs n_b unterscheidet und zudem eine Totzeit n_k existiert, ist die Bestimmung der Modellordnung ein komplexes Problem [85, S. 558].

2.2.2 Zeitdiskrete Neuronale Netze

Aus der Klasse der KNN ist das *Multilayer perceptron* (MLP) der am weitesten verbreitete Modellansatz [85, S. 241]. KNN bilden die Informationsverarbeitung biologischer Nervenzellen (Neuronen) und deren Verhalten mathematisch durch ein sog. *Perzeptron* ab. Die mathematischen Nachbildungen eines Neurons weisen eine diskrete Anzahl von Eingängen auf. Die Eingänge werden jeweils durch einen individuellen Faktor gewichtet und diese Gewichtung aufsummiert. Anschließend werden die gewichteten, aufsummierten Eingänge durch eine Aktivierungsfunktion transformiert und eine skalare Zwischengröße oder Gewichtung ausgegeben. KNN sind eine Erweiterung des mathematischen Neurons in paralleler Anordnung in einer verdeckten Schicht, verbunden mit einem Ausgangsneuron. Da KNN beliebig erweiterbare, vernetzbare Modellstruktur sind, sind sie universell einsetzbare Modelle für das Abbilden sowohl von statischen als auch dynamischen Prozessen [85, S. 241, 84, S. 64]. KNN sind in der Lage, mit entsprechend steigender Anzahl an Neuronen der verdeckten Schicht, eine jede stetige (nicht-) lineare Funktion abzubilden [54]. Ein KNN mit drei Schichten ist in Abb. 2.5 dargestellt.

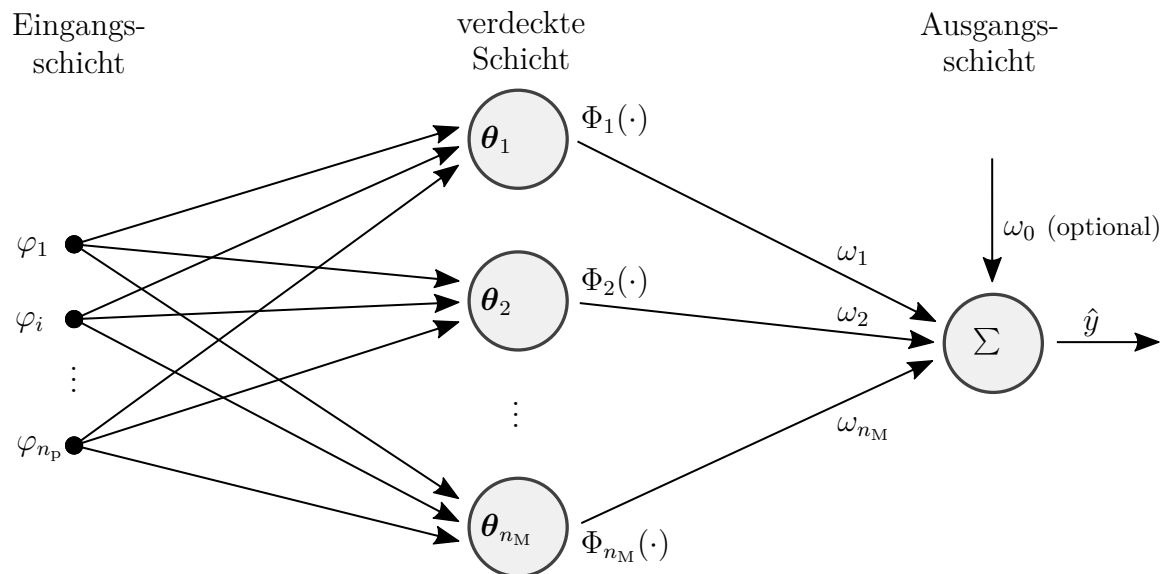


Abbildung 2.5 Aufbau einer künstlichen Neuronalen Netzstruktur (KNN) aus überlagerten Basisfunktionen Φ_i [85, S. 251, 48, S. 12]

Die Systemeingänge beschreiben im Unterschied zu den statischen Modellen die Regressoren φ anstelle der Modelleingänge \mathbf{u} . Eine KNN-Struktur lässt sich mathematisch durch $n_M + 1$ gewichtete Basisfunktionen ausdrücken [84, S. 64, 9, S. 42]

$$\hat{y}(\varphi) = w_0 + \sum_{i=0}^{n_M} \mathbf{w}_i \Phi_i(\varphi, \theta_i). \quad (2.13)$$

Der Vektor \mathbf{w}_i mit $i = [0, 1, \dots, n_M]$ beinhaltet die Gewichte der Ausgangsschicht. Zudem kann die Konstante w_0 optional addiert werden. Die Gewichte beschreiben den Beitrag des einzelnen Neurons zum Gesamtbeitrag des Modellausgangs. Als Netztopologie sind RBF- oder MLP-Netze weit verbreitet, welche sich in der Wahl der Basisfunktion unterscheiden [84, S. 65, 85, S. 241]. Ein Neuron i der verdeckten Schicht bildet einen n_p -dimensionalen Eingangsvektor $\boldsymbol{\varphi} = [\varphi_1 \ \varphi_2 \ \dots \ \varphi_{n_p}]^\top$ in Kombination mit einem Parametervektor $\boldsymbol{\theta}_i$ auf eine skalare Größe ab. Diese skalare Größe wird durch die Aktivierungsfunktion nichtlinear auf den Neuronenausgangswert transformiert [85]. Dies geschieht bei Netzen mit RBF-Topologien mittels radialer Konstruktion [48, S. 13]. Die Projektion des Eingangsvektors auf den Gewichtsvektor beschreibt bei MLP-Netzen das Skalarprodukt, welches das n_p -dimensionale Eingangsproblem kombiniert und als ein eindimensionales skalares Aktivierungsfunktionals reduziert wiedergibt. Als typische Aktivierungsfunktionen von MLP werden stetige differenzierbare Funktionen wie die Sigmoid- oder die logarithmische Funktionen gewählt [85, S. 247].

Das Skalarprodukt reduziert einen mehrdimensionalen Eingangsvektor auf eine skalare Größe. Dieser Konstruktionsmechanismus bildet Nichtlinearitäten nur in Richtung des Vektors $\boldsymbol{\theta}_i = [w_{i,0} \ w_{i,1} \ \dots \ w_{i,n_p}]^\top$ ab. In sonstige Richtungen ergibt sich ein konstantes Verhalten [84, S. 66, 48, S. 13]. Die Aktivität x_i eines Neurons kann folglich mit $\tilde{\boldsymbol{\varphi}} = [1 \ \varphi_1 \ \varphi_2 \ \dots \ \varphi_{n_p}]$ als Skalarprodukt-Konstruktion

$$x_i = \tilde{\boldsymbol{\varphi}} \cdot \boldsymbol{\theta}_i = [1 \ \varphi_1 \ \varphi_2 \ \dots \ \varphi_{n_p}] \cdot \begin{bmatrix} w_{i,0} \\ w_{i,1} \\ \vdots \\ w_{i,n_p} \end{bmatrix} \quad (2.14)$$

berechnet werden [85, S. 254, 48, S. 13]. Mit einer logarithmischen Aktivierungsfunktion kann der Ausgang eines Perzeptrons mit Gl. 2.15 bestimmt werden. Das Perzeptron hängt dabei nur von den Gewichten der verdeckten Schicht ab. Das Gewicht $w_{i,0}$ bildet ein Offset ab, der als Bias bezeichnet wird (siehe Abb. 2.6) [85, S. 211].

$$\Phi(x_i) = \tanh(x_i) \quad (2.15)$$

Das Netzwerk lässt sich um beliebig viele verdeckte Schichten erweitern. Im Rahmen dieser Arbeit wird das dreischichtige MLP verwendet. Das sog. *Three-layer-perceptron* gilt bereits als universeller Approximator für $M \rightarrow \infty$ [85, S. 251]. Die Anzahl an Parametern eines n_p -dimensionalen MLPs mit n_M Neuronen der verdeckten Schicht kann mit $n_M(n_p + 1) + n_M + 1$ ermittelt werden [85, S. 249].

Die Parameter des MLP-Netzwerkes werden meist durch eine numerische Optimierung trainiert. Verschiedene Trainingsalgorithmen sind in [88, 10] zusammengefasst. Dabei werden zwei Arten von Parametern bestimmt. Zum einen müssen die Gewichtungen der Ausgangsschicht \mathbf{w}_i trainiert werden. Diese linearen Parameter steuern die Amplituden der Basisfunktionen.

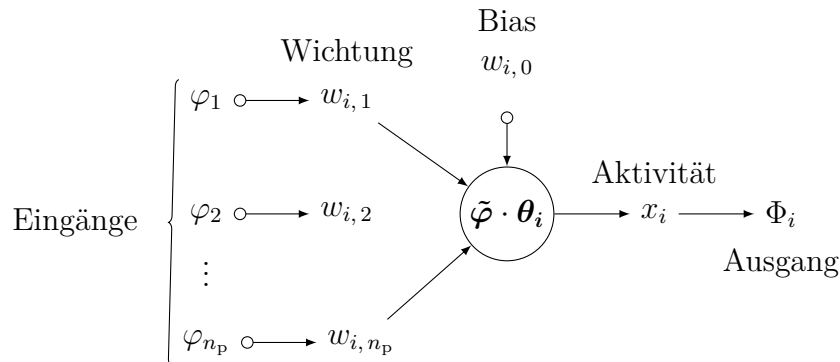


Abbildung 2.6 Ein Perzeptron: Darstellung eines Neurons i der verdeckten Schicht eines MLP [nach 85, S. 467]

Zum anderen müssen die Gewichtungen der verdeckten Schicht $w_{i,j}$ ermittelt werden. Diese nichtlinearen Parameter repräsentieren die Richtung, Steigung und die Positionen der Basisfunktionen [85, S. 252]. Die Komponenten des Regressorvektors φ hängen im Training von der Modellkonfiguration ab. Durch die Rückkopplung des Modellausgangs in Simulationsanordnung weist das KNN im Gegensatz zur Prädiktion eine rekurrente Struktur auf, da der Modellausgang $\hat{y}(k)$ als Modelleingang zum Zeitpunkt $k + 1$ dient [27, S. 38]. Ein Trainingsalgorithmus ist folglich abhängig der Topologie des Netzes zu wählen.

Als numerisches Parameterschätzverfahren ist hier der *Levenberg-Marquardt-Algorithmus* (LMA) als Optimierungsverfahren anstelle des in der modellbasierten Kalibrierung weitverbreiteten *Backpropagation*-Algorithmus ausgewählt [48, 115, 84, 85]. Dieser gradientenbasierte Algorithmus setzt eine quadratische Fehlerfunktion voraus und bietet für das Training weniger Hundert Gewichte eine geeignete Methode [27, S. 38, 8, S. 29]. Da in der Lösung der Parametervektor nichtlinear in die Kostenfunktion eingeht, führt das auf eine nichtlineare LS-Problematik zurück. Der LMA stellt dabei eine Modifikation des Gauß-Newton-Algorithmus dar und weist trotz des höheren Rechenaufwands im Vergleich zum *Backpropagation*-Verfahren ein schnelleres Konvergenzverhalten auf. Der LMA interpoliert zwischen dem Gauß-Newton-Verfahren und dem Gradientenverfahren des *steilsten Abstieges* [85, S. 105]. Ist in der Simulationsanordnung eine rekurrente Netztopologie zu trainieren, werden zusätzlich die totale Ableitung und die Ableitung nach den verzögerten Ausgängen benötigt. Die zwei am weitesten verbreiteten Trainingsalgorithmen für rekurrente Strukturen *Backpropagation through time* und *Real time recurrent learning* (RRL) sind detailliert in [85] beschrieben.

2.2.3 Anregungssignale für dynamische Systeme

Im Fall der statischen Versuchsplanung werden stationäre Betriebspunkte geplant. Diese werden am Prüfstand angesteuert und Mittelwertmessungen in diesem Betriebspunkt für die Modellbildung genutzt. Konstante Eingangssignale erzeugen keine Dynamik und sind nicht in der Lage, den Eingangsraum des nichtlinearen dynamischen Systems ausreichend anzuregen. In der dynamischen Modellierung ist eine Systemanregung für das Abbilden von Nichtlinearitäten zwingend erforderlich. Die Wahl geeigneter Anregungssignale stellt ein entscheidendes Kriterium für die erzielbare Modellgüte dar. Dabei muss ein Anregungssignal sämtliche relevanten Frequenzen des Systems und damit den gesamten Eingangsraum abbilden [85, S. 569].

Typischerweise werden zur Anregung (nicht-) periodische pseudobinäre Rauschsignale mit konstanten oder variablen Amplituden, Sinussignale oder sog. Chirp-Signale verwendet. Als solche werden Sinussignale mit zeitlich steigender Frequenz bezeichnet. Da über Prozessnichtlinearitäten selten eine Vorkenntnis existiert, ist es sinnvoll, ein universell einsetzbares Anregungssignal zu verwenden. Das *Amplitude modulated pseudorandom bit sequences* (APRBS) gehört zu den am weitesten verbreiteten Anregungssignalen für dynamische Systeme [27, 87, 52, 112, 85, 45]. Eine Systemanregung mit einem zeitkontinuierlichen periodischen Signal reicht für die Abbildung einer kompletten Systembeschreibung nicht aus [112, S. 272]. Erreicht ein konvergiertes Optimierungsverfahren eines Modells für diesen Fall dennoch eine optimale Fehlergüte, kann für die spezielle Anregung eine lokale Lösung gefunden werden. Diese besitzt weder Allgemeingültigkeit noch kann eine Aussage über die Modellgüte gegenüber höher- oder niederfrequenten Anregungssignalen getroffen werden [112, S. 272]. Folglich entspricht eine Anzahl n -gültiger lokaler Modelle einer Linearisierung des Betriebspunkts. Die Versuchsreihe in Kap. 4 beschränkt sich auf diese lokale Lösungen in einem Betriebspunkt.

Eine Diskussion gängiger Anregungssignale ist in [72, 85] zu finden. Speziell im Bereich der modellbasierten Kalibrierung für Antriebssysteme ist es von Bedeutung, eine zu starke Anregung und eine Schädigung der Aktuatoren oder des Prüflings zu vermeiden [100]. Bei der Verwendung eines APRBS sind minimale und maximale Amplitude sowie die minimale Taktzeit definierbar und so zu wählen, dass der Prüfling nicht geschädigt wird [45, S. 156]. Geeignete Ansätze in der Versuchsplanung zur Limitierung des Anregungssignals auf die wesentlichen reellen Frequenzen des dynamischen Prozesses bilden derzeit den Fokus in der Entwicklung. Eine Limitmodellierung wird beispielsweise in [62] und ein Ansatz für den Aufbau von Anregungssignalen für nichtlineare dynamische Modelle bei Verbrennungsmotoren in [6] vorgestellt.

2.3 Analytische Kennfeldglattheit

Im letzten Schritt des modellbasierten Kalibrierprozesses gilt es, die optimierten Parameter in den Skalaren, Kennlinien und Kennfeldern der Steuererätefunktionen zu hinterlegen. Ein Beispiel für ein Skalar als Führungsgröße einer Steuererätefunktion ist die Leerlaufdrehzahl. Über eine Kennlinie kann eine Korrekturfunktion etwa über die Motortemperatur abgelegt werden. Kennfelder werden in der Regel über die Kenngrößen des Motorbetriebs, der Drehzahl und einer Lastgröße aufgespannt [29, S. 5]. Diese komplexeren Steuergrößen werden über konventionelle Kennfelder approximiert, und ein Mikrocomputer interpoliert aus einer vorgegebenen Menge an Stützstellen die Funktionswerte [62, S. 154]. Die Steuererätefunktionen geben abhängig von bestimmten Ist-Werten in vordefinierten Rastern und Positionen die tabellarischen Soll-Werte vor, die beispielsweise eine Kennlinie (Vektor) oder ein Kennfeld (Matrix) abbilden. Die Kalibrierungsaufgabe besteht darin, die vordefinierten tabellarischen Raster der Kennfelder sowie Kennlinien zu befüllen, wobei die einzelnen Raster nicht zwingend äquidistant sein müssen [29, S. 15, 69, S. 68].

Die Optimierung von Kennfeldern erfolgt an Betriebspunkten, an welchen mittels Messungen Modelle von Zielgrößen gebildet werden. Die Betriebspunkte sind nicht zwangsläufig an den einzelnen Rastern, bzw. Positionen der vordefinierten Steuerfunktion vermessen. Die statistische Versuchsplanung oder eine Zyklusreduktion mittels Clusteranalyse, die einen transienten Zyklus auf stationäre Betriebspunkte abbildet, gibt die Lage der Betriebspunkte vor. Ziel der Kennfeldrechnung ist es, einen analytischen Zusammenhang zwischen den vordefinierten Positionen des Kennfeldes, den Stützstellen, und den Betriebspunkten mathematisch zu beschreiben. Dabei ist darauf zu achten, hohe Gradienten und Sprünge zu vermeiden und die Kennfelder möglichst exakt an den optimalen Werten der Betriebspunkte zu halten. Dies ist wichtig im dynamischen Betrieb wie im subjektiven Fahrverhalten und soll die Fahrbarkeit gewährleisten. Die Fahrbarkeit bedeutet harmonische Übergänge zwischen den verschiedenen realen Betriebspunkten im dynamischen Betrieb und beschreibt die subjektive Wahrnehmung des Fahrers [69, S. 65]. Bei der innermotorischen Verbrennung beeinträchtigen hohe Gradienten die Motorakustik und den Wirkungsgrad [69, 29, S. 12]. Das stellt die Kennfelderstellung vor den Zielkonflikt zwischen möglichst exakt getroffenen Punkten und einem möglichst glatten Kennfeld. Verschiedene Ansätze zur Kennfelderstellung sind aus [86, 92, 62] bekannt. Die Berechnung der Kennfeldglattheit stützt sich auf [69].

2.3.1 Bilineare Interpolation von Betriebspunkten

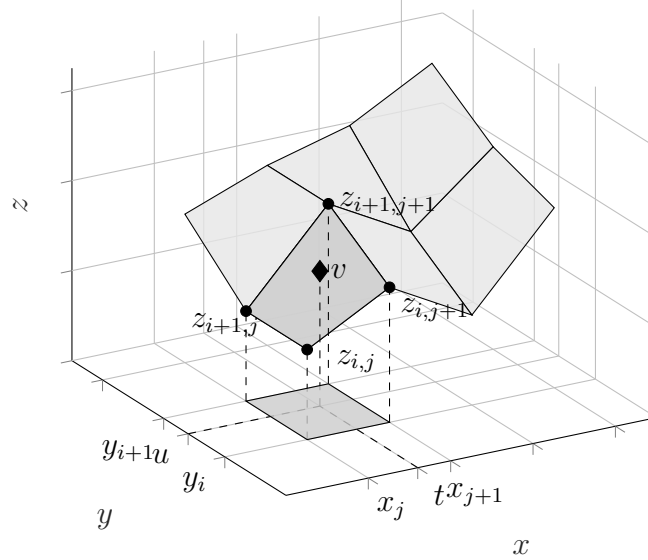


Abbildung 2.7 Bilineare Interpolation des Betriebspunkt v im Kennfeldquadranten der vier Stützstellen $[z_{i,j}, z_{i+1,j}, z_{i,j+1}, z_{i+1,j+1}]$ [angelehnt an 69, S. 65]

Dursun [29, S. 7] beweist am Beispiel des Ladedruckkennfeldes, aufgespannt über der Drehzahl und der Einspritzmenge, dass in Motorsteuergeräten die Interpolation von Sollwerten aus Kennfeldern bilinear erfolgt. Ein Kennfeld besteht aus einer Wertetabelle $z_{i,j}$ für eine bestimmte Anzahl an Stützstellen x_i und y_j . Eine analytische Beschreibung des Zusammenhangs $z_{i,j} = f(x_i, y_j)$ existiert nicht [29, S. 7].

In Abb. 2.7 ist für den Quadranten eines Kennfeldes und einen Betriebspunkt die bilineare Interpolation des entsprechenden Kennfeldes dargestellt. Ein Kennfeldwert v des Betriebspunktes an der Position $[t, u]$ kann aus den vier benachbarten Kennfeldstützstellen über die lokalen Koordinaten Gl. 2.16 ermittelt werden. Die bilineare Interpolation Gl. 2.17 berechnet durch eine Linearkombination der beteiligten Kennfeldstützstellenwerte den Kennfeldwert v des Betriebspunktes. Je nach Position $[t, u]$ beeinflusst der Betriebspunktwert v die umliegenden Kennfeldstützstellen über eine zusammengefasste Gewichtung \mathbf{m} [69, S. 65, 29, S. 8].

$$t = \frac{x - x_j}{x_{j+1} - x_j} \quad \text{und} \quad u = \frac{y - y_i}{y_{i+1} - y_i} \quad (2.16)$$

$$v = \underbrace{(1-t)(1-u)}_{m_{i,j}} z_{i,j} + \underbrace{t(1-u)}_{m_{i,j+1}} z_{i,j+1} + \underbrace{u(1-t)}_{m_{i+1,j}} z_{i+1,j} + \underbrace{tu}_{m_{i+1,j+1}} z_{i+1,j+1} \quad (2.17)$$

Die Gl. 2.17 kann wiederum als Vektormultiplikation ausgedrückt werden. Mit den Gewichtungsfaktoren \mathbf{m} und den umliegenden Kennfeldstützstellen in Matrixschreibweise ergibt sich Gl. 2.18 [29, S. 8]. Die vektorielle Form der bilinearen Interpolation kann auf eine beliebige Anzahl Betriebspunktwerte mit unterschiedlichen Positionen im Kennfeld erweitert werden.

Die mehrdimensionale Interpolation erweitert sich für einen Vektor \mathbf{v} von N_{BP} Betriebspunkten, wobei die jeweiligen Gewichtungen \mathbf{m} in einer Interpolationsmatrix \mathbf{M} zusammengefasst werden. Die Größe der Interpolationsmatrix \mathbf{M} ergibt sich aus der Anzahl an Betriebspunkten N_{BP} sowie der Anzahl an Kennfeldstützstellen $N_{\text{KP}} = N_x \cdot N_y$ zu $N_{\text{BP}} \cdot N_{\text{KP}}$. Die Erweiterung auf eine beliebige Anzahl Betriebspunkte überführt Gl. 2.18 auf das lineare Gleichungssystem 2.19 [29, S. 9].

$$\mathbf{v} = \underbrace{\begin{bmatrix} m_{i,j} & m_{i,j+1} & m_{i+1,j} & m_{i+1,j+1} \end{bmatrix}}_{\mathbf{m}} \cdot \underbrace{\begin{bmatrix} z_{i,j} \\ z_{i,j+1} \\ z_{i+1,j} \\ z_{i+1,j+1} \end{bmatrix}}_{\mathbf{z}} \quad (2.18)$$

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \vdots \\ \mathbf{m}_N \end{bmatrix} \cdot \mathbf{z} = \mathbf{M} \cdot \mathbf{z}. \quad (2.19)$$

Im Umkehrschluss können für gegebene Betriebspunkte \mathbf{v} die tabellarischen Sollwerte \mathbf{z} einer Steuergerätefunktion mithilfe einer Fehlerfunktion berechnet werden. Eine nicht eindeutige Lösung ergibt sich für den Extremfall in Abb. 2.8a. Die Kennlinie zeigt ein Beispiel für einen mehrdeutig bestimmbar Stützstellenwert. In dem Fall ist der Rang der Matrix (\mathbf{M}) kleiner als die Anzahl ihrer Spalten N_{KP} und folglich kleiner als die Anzahl zu bestimmender Kennfeldstützstellenwerte [29, S. 11]. Dieser unterbestimmte Fall als zusätzlicher Freiheitsgrad bedeutet eine beliebige Lösung bei der Berechnung eines Kennfeldes. In Abb. 2.8 sind Beispiellösungen für die Erstellung von Kennlinien an den jeweiligen Stützstellen für bestimmte Betriebspunkte gezeigt. Eindeutige Lösungen dieser Problemstellung werden mithilfe des Kriteriums der Kennfeldglattheit realisiert.

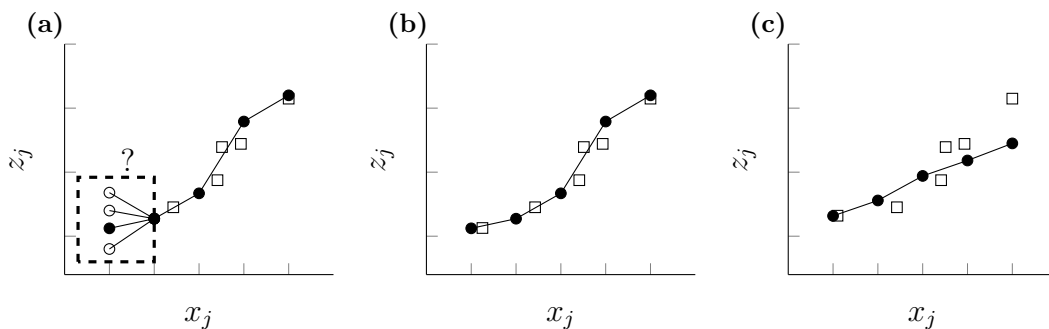


Abbildung 2.8 Schematische Darstellung einer Kennlinie des Problems mehrdeutiger Lösungen (a), mit Gewichtung zu einer exakt approximierten Lösung (b) sowie einer möglichst glatten Lösung (c) zwischen Messungen (□) und Kennfeldstützstelle (●)

2.3.2 Glattheitsmaß für Kennfeldstützstellen

Die Glattheit eines Kennfeldes wird über dessen Steigung und Krümmung der jeweiligen Achsen in x - und y -Richtung definiert. Bei der Erstellung einer Kennlinie oder eines Kennfeldes ist ein Kompromiss zwischen Exaktheit (Abb. 2.8b) und Glattheit (Abb. 2.8c) der Lösung zu finden. Da der funktionale Zusammenhang $z = f(x, y)$ eines Kennfeldes diskretisiert vorliegt, werden für die Ableitungen Differenzen verwendet. Die Steigung des Kennfeldes in x - und y -Richtung wird mit den Vorwärtsdifferenzen erster Ordnung in Abhängigkeit von zwei Punkten in Gl. 2.20 beschrieben [69, S. 66, 29, S. 13]. Die zweite Ableitung beschreibt die Krümmung des Kennfeldes. Diese wird mittels Differenzenbildung zweier aufeinander folgender erster Ableitungen und Division mit der abgedeckten Stützstellenstrecke gebildet. Die Vorwärtsdifferenzen erster und zweiter Ordnung sind für ein Kennfeld mit den Gl. 2.20 bis 2.22 formuliert [69, S. 66, 29, S. 13].

$$\frac{\Delta z_{i,j}}{\Delta x_j} = \frac{z_{i,j+1} - z_{i,j}}{x_{j+1} - x_j} \quad \text{und} \quad \frac{\Delta z_{i,j}}{\Delta y_i} = \frac{z_{i+1,j} - z_{i,j}}{y_{i+1} - y_i} \quad (2.20)$$

$$\frac{\Delta^2 z_{i,j}}{\Delta x_j^2} = \frac{2}{x_{j+2} - x_j} \left(\frac{z_{i,j+2} - z_{i,j+1}}{x_{j+2} - x_{j+1}} - \frac{z_{i,j+1} - z_{i,j}}{x_{j+1} - x_j} \right) \quad (2.21)$$

$$\frac{\Delta^2 z_{i,j}}{\Delta y_i^2} = \frac{2}{y_{i+2} - y_i} \left(\frac{z_{i+2,j} - z_{i+1,j}}{y_{i+2} - y_{i+1}} - \frac{z_{i+1,j} - z_{i,j}}{y_{i+1} - y_i} \right) \quad (2.22)$$

Für die Bildung einer Kostenfunktion mit Glattheitsmaß werden die einzelnen Beiträge der Gl. 2.20 bis 2.22 quadriert. Das Gleichungssystem in Vektorschreibweise ergibt sich bei Aufteilung der Beiträge in einen Vektor der Kennfeldstützstellen \mathbf{z} und einer Matrix \mathbf{R} . Die Glattheitskosten der Steigung und Krümmung aus erster und zweiter Ableitung sind definiert in den folgenden vier Gleichungen für x - und y -Richtung [69, S. 67, 29, S. 14]:

$$c_{\text{Rau},x} = \sum_i^{N_y} \sum_j^{N_x-1} \left(\frac{\Delta z_{i,j}}{\Delta x_j} \right)^2 = \mathbf{z}^\top \mathbf{R}_x^\top \mathbf{R}_x \mathbf{z} \quad (2.23)$$

$$c_{\text{Rau},y} = \sum_i^{N_y-1} \sum_j^{N_x} \left(\frac{\Delta z_{i,j}}{\Delta y_i} \right)^2 = \mathbf{z}^\top \mathbf{R}_y^\top \mathbf{R}_y \mathbf{z} \quad (2.24)$$

$$c_{\text{Rau},xx} = \sum_i^{N_y} \sum_j^{N_x-2} \left(\frac{\Delta^2 z_{i,j}}{\Delta x_j^2} \right)^2 = \mathbf{z}^\top \mathbf{R}_{xx}^\top \mathbf{R}_{xx} \mathbf{z} \quad (2.25)$$

$$c_{\text{Rau},yy} = \sum_i^{N_y-2} \sum_j^{N_x} \left(\frac{\Delta^2 z_{i,j}}{\Delta y_i^2} \right)^2 = \mathbf{z}^\top \mathbf{R}_{yy}^\top \mathbf{R}_{yy} \mathbf{z} \quad (2.26)$$

Die Matrix \mathbf{R} enthält Informationen über die Lage der Stützstelle im Kennfeld. Die Matrix \mathbf{R}_x enthält Einträge aus genau zwei Kennfeldeinträgen der umliegenden Stützstellen. Eine Multiplikation der Kennfeldstützstellen mit der Matrix \mathbf{R}_x entspricht der vektoriellen Darstellung des Gradienten oder der Steigung in x -Richtung.

Die Matrix \mathbf{R}_{xx} besteht aus Termen dreier umliegender Stützstellenpositionen und beschreibt die Kennfeldkrümmung in x -Richtung [69, S. 67, 29, S. 14].

Eine vollständige Kostenfunktion für das Kriterium der Kennfeldglattheit ergibt sich mit einer Gewichtung der einzelnen Glattheitsbeiträge aus den Gl. 2.23 bis 2.26, siehe Gl. 2.27. Die vier Gewichtungen ζ_i erlauben die Einstellung der Steigung oder Krümmung in beide Kennfeldrichtungen. Das Gütekriterium der Glattheit ergibt sich durch die Substitution einer Glattheitsmatrix \mathbf{R}_d in Gl. 2.29 [29, S. 15].

$$c_{\text{Rau}}(\mathbf{z}) = \zeta_x c_{\text{Rau},x} + \zeta_y c_{\text{Rau},y} + \zeta_{xx} c_{\text{Rau},xx} + \zeta_{yy} c_{\text{Rau},yy} \quad (2.27)$$

$$c_{\text{Rau}}(\mathbf{z}) = \mathbf{z}^\top \underbrace{(\zeta_x \mathbf{R}_x^\top \mathbf{R}_x + \zeta_y \mathbf{R}_y^\top \mathbf{R}_y + \zeta_{xx} \mathbf{R}_{xx}^\top \mathbf{R}_{xx} + \zeta_{yy} \mathbf{R}_{yy}^\top \mathbf{R}_{yy})}_{\mathbf{R}_d} \mathbf{z} \quad (2.28)$$

$$c_{\text{Rau}}(\mathbf{z}) = \mathbf{z}^\top \mathbf{R}_d \mathbf{z} \quad (2.29)$$

Die Ableitung der Gl. 2.29 nach den einzelnen Kennfeldkomponenten \mathbf{z} ergibt eine lineare Gleichung in \mathbf{z} mit

$$\frac{\partial c_{\text{Rau}}}{\partial \mathbf{z}} = 2\mathbf{R}_d \mathbf{z} = 0. \quad (2.30)$$

Da nur die Glattheit berücksichtigt ist, erfolgt die glatte Lösung in Form eines ebenen Kennfeldes mit $\mathbf{z} = 0$. Folglich fehlt ein Betrag, der die Kennfeldwerte an exakten Punkten festsetzt [69, S. 69].

2.3.3 Kennfeldglattheit unter Vorgabe exakter Punkte

Sollen Betriebspunkte durch die benachbarten Stützstellen exakt abgebildet werden, erweitert sich die Anforderung an eine Kostenfunktion der Kennfelderstellung um eine zusätzliche Nebenbedingung. Eine Formulierung des Optimierungsproblems für eine analytische Berechnung ist die glatteste Kennfeldlösung unter der Berücksichtigung von exakten Festwerten an bestimmten Positionen im Kennfeld. Die exakt einzuhaltenden Zahlenwerte von Betriebspunkten \mathbf{v}_{ex} mit der zugehörigen Interpolationsmatrix \mathbf{M}_{ex} lassen sich analog zu Gl. 2.19 mit dem folgenden linearen Gleichungssystem vorgeben [69, S. 70, 29, S. 17]

$$\mathbf{v}_{\text{ex}} = \mathbf{M}_{\text{ex}} \mathbf{z}. \quad (2.31)$$

Die Interpolationsmatrix \mathbf{M}_{ex} beschreibt die Interpolationsvorschrift in Abhängigkeit von den Punkten $[\mathbf{u}_{\text{ex}}, \mathbf{t}_{\text{ex}}]$ des Vektors einzuhaltender Werte \mathbf{v}_{ex} . Der Vektor \mathbf{z} beinhaltet das mittels der bilinearen Interpolation zu approximierende Kennfeld. Unter der Nutzung des Lagrange-Verfahrens kann die lineare Gleichheitsbedingung in Gl. 2.31 als Nebenbedingung des Optimierungsproblems formuliert werden [69, S. 70, 29, S. 16]:

$$r(\mathbf{z}) = \mathbf{M}_{\text{ex}} \mathbf{z} - \mathbf{v}_{\text{ex}} = 0 \quad (2.32)$$

Jeder Festwert eines Betriebspunktes \mathbf{v}_{ex} wird entsprechend dem Kennfeldquadrat von maximal vier Kennfeldstützstellen aufgespannt. Die Formulierung der Kennfeldglattheit Gl. 2.29 wird um die Randbedingung aus Gl. 2.32 erweitert. Zur Lösung wird das Lagrange-Verfahren für das beschränkte Optimierungsproblem verwendet [69, S. 70, 29, S. 17]. Die Randbedingung in Gl. 2.32 wird mit einem Vektor aus Lagrange-Multiplikatoren $\boldsymbol{\lambda}$ multipliziert. Somit ergibt sich ein skalarer Wert. Die Gesamtkosten ergeben sich zu

$$\mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}) = \mathbf{z}^\top \mathbf{R}_d \mathbf{z} + \boldsymbol{\lambda}^\top (\mathbf{M}_{\text{ex}} \mathbf{z} - \mathbf{v}_{\text{ex}}), \quad (2.33)$$

in Abhängigkeit von \mathbf{z} und $\boldsymbol{\lambda}$ [69, S. 70, 29, S. 17]. Das Lagrange-Verfahren reduziert das Nebenbedingungsproblem auf ein beschränktes Optimierungsproblem. Dieses Optimierungsproblem wird durch Minimierung der Lagrange-Funktion Gl. 2.33 mittels Ableitung nach \mathbf{z} und $\boldsymbol{\lambda}$ gelöst [29, S. 17]. Die Ableitung und das Gleichsetzen mit null $\nabla_{\mathbf{z}, \boldsymbol{\lambda}} \mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}) = \mathbf{0}$ führt auf das invertierte Gleichungssystem in Gl. 2.34. Als Bedingung für dessen Gültigkeit muss die Position eines jeden Betriebspunktes innerhalb des Kennfeldbereiches liegen [69, S. 70]. Sind mehrere Betriebspunkte in einem Kennfeldquadranten exakt zu treffen, ist das Gleichungssystem nicht lösbar. Die Gl. 2.31 ist für diesen Fall überbestimmt.

$$\begin{bmatrix} \hat{\mathbf{z}} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} 2\mathbf{R}_d & \mathbf{M}_{\text{ex}}^\top \\ \mathbf{M}_{\text{ex}} & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{v}_{\text{ex}} \end{bmatrix} \quad (2.34)$$

Mit der analytischen Kennfeldglattheit ist es möglich, die Anzahl der Variablen für das Optimierungsproblem der Kennfelderstellung stark zu reduzieren [69, S. 71]. Im Anwendungsbeispiel Kap. 4 in der transienten Untersuchung wird der hier vorgeschlagene Ansatz zur Kennfelderstellung bereits in der Versuchsplanung verwendet. Aus wenigen Variationspunkten werden jeweils Kennfelder aufgespannt. Die Variationspunkte bilden die exakt abzubildenden Punkte des zu approximierenden Kennfeldes ab. Diese werden durch eine statistische Versuchsplanung im begrenzten Variationsbereich an festgelegten Positionen im Kennfeld variiert und für jede Variation ein Kennfeld berechnet. Der Vorteil liegt hierbei in der Reduzierung der Optimierungsgrößen im Vergleich zu einer Vermessung sämtlicher Kennfeldwerte. Mit einer geeigneten Verteilung der Variationspunkte im Kennfeld können mit dem analytischen Kennfeldansatz möglichst glatte Kennfelder für die transiente Untersuchung am Prüfstand erstellt werden. Dabei sind die Glattheitskosten der Kennfeldlösung aus den Aspekten der Fahrbarkeit sowie der Sicherheit der Stellglieder bei Überbeanspruchung im dynamischen Betrieb skalierbar.

3 Beiträge zur modellbasierten Kalibrierung

Im Vergleich zur stationären Kalibrierung von Antriebssystemen erhöht sich die Anzahl an Freiheitsgraden bei transienten Aufgaben. Daher werden zur Eingrenzung des hohen Kalibrieraufwandes Sensitivitätsmethoden vorgeschlagen, mit denen eine modellbasierte Problemstellung auf die wesentlichen Einflussfaktoren reduziert wird. Die Sensitivitätsanalyse bewertet den Einfluss vieler Eingangsparameter auf die Modellausgänge qualitativ und hilft ein Systemverständnis aufzubauen. Im weiteren Verlauf der Arbeit wird der Begriff der Sensitivitätsanalyse in die Methoden der Datenanalyse und der Sensitivitätsanalyse aufgeteilt. Die Sensitivitätsanalyse setzt im Rahmen dieser Arbeit eine Modellbildung bei der Analyse der Sensitivität voraus. Die Methoden der Datenanalyse beschreiben die statistischen Zusammenhänge der Daten untereinander, anhand derer die Sensitivität ermittelt wird.

Die transiente Analyse am Prüfstand ist durch beispielsweise Totzeiten beim Transport von Masse in Rohrleitungen oder durch verschiedene Messsysteme, die Messgrößen nicht zeitsynchron aufzeichnen, beeinflusst. Im dynamischen Motorbetrieb bewirken die Totzeiten im System selber Verzugszeiten der Stellglieder und folglich Prozessverzögerungen [32]. Der Einfluss der Totzeit kann mit einer genauen Schätzung minimiert werden. Daher werden verschiedene Schätzalgorithmen vorgestellt und verglichen.

3.1 Methoden der Datenanalyse

Im folgenden Unterkapitel ist der Korrelationskoeffizient dargestellt sowie der Rangkorrelationskoeffizient. Die Korrelationskoeffizienten berechnen ein Maß der linearen oder monotonen Beziehung zweier Signale. Ergänzend wird ein Maß aus der Informationstheorie vorgestellt, mit dem sich auch nichtlineare Zusammenhänge in Messdaten erkennen lassen.

3.1.1 Empirische Korrelationskoeffizienten

Die Korrelation drückt die statistische Beziehung zweier Signale zueinander aus. Eine lineare Abhängigkeit zwischen den Signalen $u(k)$ und $y(k)$ kann durch das Verhältnis der geschätzten Kovarianz s_{uy} und dem Produkt der geschätzten Standardabweichungen $s_u s_y$ ausgedrückt werden [26, S. 8]. Das lineare Maß der Korrelation bildet einen Wert im Wertebereich $[-1, 1]$ ab. Der Zusammenhang nach *Pearson* kann mit den Mittelwerten der beiden Signale \bar{u} und \bar{y} anhand

$$r_P = \frac{s_{uy}}{s_u s_y} = \frac{\sum_{k=1}^N (u_k - \bar{u})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^N (u_k - \bar{u})^2} \sqrt{\sum_{k=1}^N (y_k - \bar{y})^2}} \quad (3.1)$$

beschrieben werden [26, S. 8]. Ein hoher statistischer Zusammenhang ergibt Werte nahe -1 und 1 . Trifft die Korrelation auf einen dieser Werte, sind die Signale identisch (1) bzw. gegenphasig (-1) und liegen genau auf einer Linie. Kleinere Werte bedeuten einen niedrigeren linearen statistischen Zusammenhang. Stochastische Unabhängigkeit bedeutet der Wert $r_P = 0$. Von einer *schwachen Korrelation* kann für den Bereich $0 < |r| < 0,5$, von einer *starken Korrelation* kann für $0,8 \leq |r| < 1$ ausgegangen werden [23, S. 111].

Der Korrelationskoeffizient nach *Pearson* ist einzig für lineare Beziehungen der beiden Variablen u und y gültig [26, S. 8]. Daher wird zudem der Rangkorrelationskoeffizient nach *Spearman* eingeführt. Dieser Korrelationskoeffizient betrachtet im Vergleich zu *Pearson* nicht die Werte der Stichprobe, sondern die Reihenfolge der Messdatenwerte bzw. deren Rangwert. Der Rangkorrelationskoeffizient nach *Spearman*

$$\rho_{SP} = \frac{\sum_{k=1}^N (R(u_k) - \overline{R(u)})(R(y_k) - \overline{R(y)})}{\sqrt{\sum_{k=1}^N (R(u_k) - \overline{R(u)})^2} \sqrt{\sum_{k=1}^N (R(y_k) - \overline{R(y)})^2}} \quad (3.2)$$

liefert Ergebnisse für monotone Signale und ist somit weniger empfindlich gegenüber Ausreißern als der Korrelationskoeffizient nach *Pearson*. Der Rangkorrelationskoeffizient ist mit der rangtransformierten Zuordnung $R(u_k)$ der Signalgröße $u(k)$ und dem entsprechenden Mittelwert der Rangfolge $\overline{R(u)}$ gegeben [102, S. 203]. Die Rangtransformation ist detailliert inklusive Beispiel und der Bedeutung mehrfacher identischer numerischer Werte innerhalb einer Zeitreihe in Kap. B.3 im Anhang dargestellt.

3.1.2 Verallgemeinerte Transinformation

In der Informationstheorie wird der Informationsgehalt von Messdaten mithilfe der Shannon-Entropie quantifiziert [114]. Die informationstheoretischen Begriffe bilden die Grundlage für die Beschreibung der Transinformation, welche ein Maß für den Informationsfluss zwischen Ein- und Ausgang eines beliebigen Systems darstellt. Zur Beschreibung der Transinformation werden die Begriffe Informationsgehalt und Entropie aus informationstheoretischer Betrachtungsweise als Informationsmaße definiert [114]. In der Literatur existieren zahlreiche Definitionen für den Begriff der Information. Gebräuchlicher Weise werden Information und Entropie synonym als Informationsmaß bezeichnet [89, S. 83]. Der Informationsgehalt I_i des Zeichens $u_i \in \mathbf{u}$ mit der Auftrittswahrscheinlichkeit p_i ist

$$I_i = -\text{ld}(p_i) \quad \text{mit} \quad [I] = \text{bit}. \quad (3.3)$$

Die Wahrscheinlichkeit p_i und der Informationsgehalt I_i stehen in umgekehrt proportionalen Verhältnis. Im Allgemeinen verwendet die Informationstheorie den *logarithmus dualis* (Logarithmus zur Basis 2) mit der Einheit *bit* zur Angabe des Informationsgehalts [114]. Die Gl. 3.3 liefert bei einem häufig auftretenden Zeichen, also einem wahrscheinlichen Ereignis, einen geringen Informationsgehalt. Ein sicheres Ereignis mit der Wahrscheinlichkeit $p_i = 1$ ergibt einen Informationsgehalt von null. Umgekehrt ergeben selten auftretende Zeichen einen hohen Informationsgehalt. Ein unmögliches Ereignis mit $p_i = 0$ lässt I_i gegen unendlich streben [126, S. 4].

Die informationstheoretische Entropie beschreibt den mittleren Informationsgehalt einer gedächtnislosen Quelle \mathbf{U} mit dem Zeichenvorrat $\mathbf{U} = \{u_1, u_2, \dots, u_n\}$ und den Wahrscheinlichkeiten $p_U(\mathbf{U})$, siehe Gl. 3.4. Die Entropie kann als Erwartungswert des Informationsgehaltes verstanden werden und misst die Ungewissheit, die im Mittel durch die Zeichen der Quelle aufgelöst wird [126, S. 5]. Die Entropie $H(\mathbf{U})$ und $H(\mathbf{Y})$ der Zufallsvariablen \mathbf{U} und \mathbf{Y} sowie die gemeinsame Verbundentropie $H(\mathbf{U}, \mathbf{Y})$ für das Auftreten paarweiser Zeichen lassen sich wie folgt definieren [114]:

$$H(\mathbf{U}) = - \sum_{i=1}^N p_U(u) \text{ld}(p_U(u)) \quad (3.4)$$

$$H(\mathbf{Y}) = - \sum_{j=1}^M p_Y(y) \text{ld}(p_Y(y)) \quad (3.5)$$

$$H(\mathbf{U}, \mathbf{Y}) = - \sum_{i=1}^N \sum_{j=1}^M p_{UY}(u, y) \text{ld}(p_{UY}(u, y)) \quad (3.6)$$

Die Verbindung des dynamischen Informationsbegriffes mit dem der Thermodynamik ist in [89, 22] beschrieben. Veranschaulichen lässt sich das Prinzip von Shannons Entropie und der Transinformation mit dem theoretischen Modell eines Informationskanals [22].

Der Informationskanal beschreibt die Übertragung zwischen einem Sender und einem Empfänger und ist in Abb. 3.1 dargestellt. Die Transinformation misst denjenigen Teil der vom Sender gesendeten Information, der beim Empfänger ankommt. Bei einem idealen, störungsfreien Informationskanal empfängt der Empfänger die gesamte Information des Senders und die Transinformation ist maximal. Werden Messdaten betrachtet, werden deren Signalwerte als Zufallsvariablen angesehen. Der Informationskanal in Abb. 3.1 beschreibt die beiden Informationsgehalte der Zufallsvariablen des Eingangs \mathbf{U} und des Ausgangs \mathbf{Y} , sowie exemplarisch deren Verlust und Störung. Wenn der Verlust und die Störung des Kanals so groß sind, dass die Transinformation verschwindet, dann sind die beiden Zufallsgrößen statistisch unabhängig. [22] definiert das Maß der Transinformation als den Verlust an Unsicherheit im Mittel durch die Kenntnis des Ergebnisses einer anderen Zufallsvariablen und interpretiert diese als die Kommunikationsrate des Kanals in Anwesenheit von Rauschen und Störungen.

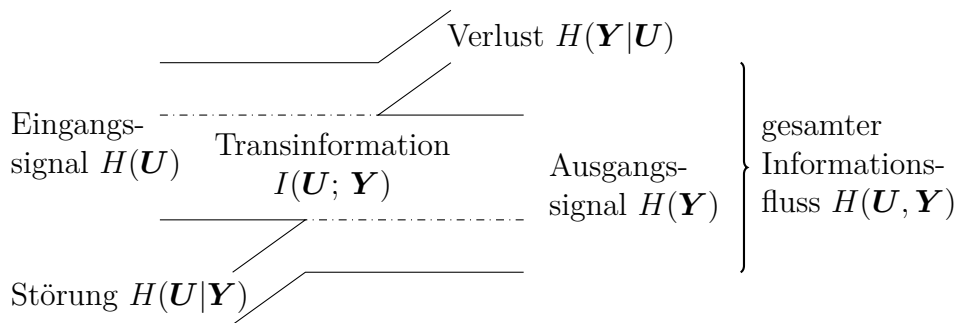


Abbildung 3.1 Der Informationskanal nach [114] und die Verdeutlichung der Zusammenhänge und Bedeutungen der Entropien als Informationsmaße [nach 38, S. 29]

Die Entropien sind jeweils abhängig von der Wahrscheinlichkeitsfunktion $p_U(u)$, $p_Y(y)$ oder der Verbundwahrscheinlichkeitsfunktion $p_{UY}(u,y)$. Die bedingte Entropie $H(\mathbf{Y}|\mathbf{U})$ misst die Unsicherheit der Variable \mathbf{U} bei der Kenntnis einer Zufallsvariablen \mathbf{Y} . Die bedingte Entropie ergibt sich aus der Kettenregel zu $H(\mathbf{Y}|\mathbf{U}) = H(\mathbf{U}, \mathbf{Y}) - H(\mathbf{U})$ [22, S. 22]. Die wechselseitige Information bzw. Transinformation $I(\mathbf{U}; \mathbf{Y})$ kann durch die einzelnen Informationsmaße ausgedrückt werden (siehe Abb. 3.1) und lässt sich aus dem Zusammenhang der Entropien bestimmen:

$$I(\mathbf{U}; \mathbf{Y}) = H(\mathbf{U}) + [H(\mathbf{Y}) - H(\mathbf{U}, \mathbf{Y})] \quad (3.7)$$

Die Transinformation misst den statistischen Zusammenhang zweier Zufallsvariablen. Die Herausforderung bei der Bestimmung der Transinformation liegt in der

Schätzung der Wahrscheinlichkeitsverteilungen $p_U(u)$, $p_Y(y)$ und $p_{UY}(u,y)$, besonders im Fall experimenteller Daten [94, S. 73].

In der Literatur ist die Wahrscheinlichkeitsdichteschätzung mittels histogrammbasierten Methoden, die Kerneldichteschätzung und die Schätzung von Wahrscheinlichkeitsdichteverhältnissen mit der LS-Methode bekannt [104, 40, 119, 33, 51]. In dieser Arbeit erfolgt die Schätzung der jeweiligen Entropien und der Transinformation über eine verallgemeinerte Transinformation und die Implementierung [nach 38, S. 39ff, 94, 93]. Dieser Ansatz verspricht numerische Vorteile in der Schätzung der Informationsmaße aufgrund folgender Annahmen:

1. Der Datensatz \mathbf{Y} wird gleichverteilt. Dazu werden die Ausgangsdaten mit $t : u(k) \rightarrow R(k)$ in Gl. B.15 im Anhang transformiert und zur relativen Rangfolge im Bereich $r(k) \in [0, 1]$ normiert.
2. Zur Berechnung wird anstelle der Shannon-Entropie die Renyi-Entropie der zweiten Ordnung verwendet.

Die Renyi-Entropie H_α mit der Ordnung α und die Renyi-Information sind in folgenden Gleichungen definiert [89, S. 84]:

$$H_\alpha(\mathbf{U}) = \begin{cases} \frac{1}{1-\alpha} \text{ld} \sum_{i=1}^N p_U^\alpha(u) & ; \quad \alpha \geq 0, \alpha \neq 1 \\ -\sum_{i=1}^N p_U(u) \text{ld} p_U(u) & ; \quad \alpha = 1 \end{cases} \quad (3.8)$$

$$I_\alpha(\mathbf{U}; \mathbf{Y}) = H_\alpha(\mathbf{U}) + [H_\alpha(\mathbf{Y}) - H_\alpha(\mathbf{U}, \mathbf{Y})] \quad (3.9)$$

Für den Fall, dass $\alpha = 0$ ist, erhält man die Hartley-Entropie [89, S. 84]. Die Ordnung $\alpha = 1$ der Renyi-Entropie entspricht der Shannon-Entropie [89, S. 84, 38, S. 31]. Shannons Informationsmaße liefern für eine beliebige, z. B. eindimensionale Verteilung eine nicht negative gegenseitige Information. Folgt für die gegenseitige Information der Wert null, existiert keine statistische Abhängigkeit [93]. Das Informationsmaß nach Shannon $I_1(\mathbf{U}; \mathbf{Y})$ und die auf der Korrelationsentropie basierende verallgemeinerte Transinformation $I_2(\mathbf{U}; \mathbf{Y})$ sind definiert als [38, S. 34]:

$$I_1(\mathbf{U}; \mathbf{Y}) = \sum_{i=1}^N \sum_{j=1}^M p_{UY}(u, y) \text{ld} \frac{p_{UY}(u, y)}{p_U(u) p_Y(y)}, \quad (3.10)$$

$$I_2(\mathbf{U}; \mathbf{Y}) = \text{ld} \frac{\sum_{i=1}^N \sum_{j=1}^M p_{UY}(u, y)}{\sum_{i=1}^N p_U(u) \sum_{j=1}^M p_Y(y)}. \quad (3.11)$$

Für $\alpha = 2$ beschreibt die Renyi-Entropie die Korrelationsentropie und damit die verallgemeinerte Transinformation. Die erforderlichen Informationsmaße von Renyi werden mit dem *Grassberger-Procaccia-Takens-Algorithmus* geschätzt [93]. Auf diesem basiert die Implementierung in *Matlab* von [38], welche im Rahmen dieser Arbeit

genutzt wird. Das Informationsmaß $I_2(\mathbf{U}; \mathbf{Y})$ besitzt im Falle gleichverteilter Datenmengen die gleichen Eigenschaften für unterschiedliche Korrelationen in Messdaten wie I_1 (vgl. den Beweis [von 38, S. 34]).

Beide gegenseitigen Informationen I_1 und I_2 können als nichtlineares Pendant zur Korrelationsfunktion betrachtet werden. Zeigt ein Korrelationskoeffizient keine linearen statistischen Abhängigkeiten an, sind die von I_1 und I_2 erkannten Abhängigkeiten nichtlinear. Ein direkter Vergleich der Zahlenwerte ist unzulässig [94, S. 76]. Für eine beliebige Verteilung von \mathbf{U} und \mathbf{Y} gilt ein Wert für die Größe $0 \leq I_\alpha$ nur für $\alpha = 0$ oder $\alpha = 1$ [94, S. 75]. Daher ist eine Rangtransformation in relative Rangzahlen und somit eine Gleichverteilung im Bereich $r(k) \in [0, 1]$ notwendig. Für diesen Fall ist die verallgemeinerte Transinformation I_2 nicht-negativ und gleich null, wenn keine linearen und nichtlinearen statistischen Abhängigkeiten bestehen. Von einer verschwindenden verallgemeinerten Transinformation I_2 lässt sich nicht auf eine statistische Unabhängigkeit schließen, wenn die Verteilung von \mathbf{Y} nicht gleich wahrscheinlich ist [94, S. 75].

Der Schätzwert der normierten verallgemeinerten Transinformation berechnet sich mit der Gl. 3.12 [38, S. 42]. Diese Definition ergibt sich durch die Eigenschaft der Beschränktheit der Transinformation $0 \leq I(\mathbf{U}; \mathbf{Y}) \leq \min\{H(\mathbf{U}), H(\mathbf{Y})\}$ (siehe Tab. 3.1).

$$\hat{I}_2(\mathbf{U}; \mathbf{Y}) = \frac{H_2(\mathbf{U}) + H_2(\mathbf{Y}) - H_2(\mathbf{U}, \mathbf{Y})}{\min\{H_2(\mathbf{U}), H_2(\mathbf{Y})\}} \quad (3.12)$$

Alternative Normierungen bzw. Sensitivitätsmaße der Transinformation sind enthalten in [74, 24, 95, S. 761, 39, 75]. Die Eigenschaften der Transinformation sind für die Zufallsvariablen \mathbf{U} und \mathbf{Y} in Tab. 3.1 zusammengefasst [94, S. 77].

Tabelle 3.1 Eigenschaften der verallgemeinerten Transinformation \hat{I}_2 für \mathbf{U} und \mathbf{Y} bei einer Gleichverteilung von \mathbf{Y} [nach 38, S. 31, 94, S. 75].

1	Symmetrie	$\hat{I}_2(\mathbf{U}; \mathbf{Y}) = I_2(\mathbf{Y}; \mathbf{U})$
2	Beschränktheit	$0 \leq \hat{I}_2(\mathbf{U}; \mathbf{Y}) \leq \min\{H_2(\mathbf{U}), H_2(\mathbf{Y})\}$
3	Unabhängigkeit	$\hat{I}_2(\mathbf{U}; \mathbf{Y}) = 0 \Leftrightarrow \mathbf{X}$ und \mathbf{Y} sind unabhängig
4	Determiniertheit	$\hat{I}_2(\mathbf{U}; \mathbf{Y}) = H_2(\mathbf{Y}) \Leftrightarrow \mathbf{Y}$ eine Funktion von \mathbf{X} ist
		$\hat{I}_2(\mathbf{U}; \mathbf{Y}) = H_2(\mathbf{U}) \Leftrightarrow \mathbf{U}$ eine Funktion von \mathbf{Y} ist

3.1.3 Anwendungsbeispiele der Datenanalyse

Nachfolgend werden drei Anwendungen zur Datenanalyse vorgestellt. Sie behandeln den Einfluss einer Rangtransformation auf das Signalspektrum sowie den statistischen Zusammenhang von linear und nichtlinear abhängigen Zufallsdaten mit den

zuvor eingeführten Kenngrößen. Ein Ansatz zur Bestimmung einer geeigneten Abtastzeit für die dynamische Modellbildung ist mit der Verbundentropie vorgestellt.

Rangtransformation von Signaldaten

Für das indizierte Moment M einer urbanen RDE-Testsequenz mit Start-Stopp-Anteilen ist das Resultat der Rangtransformation in Abb. 3.2 exemplarisch dargestellt. Ein Maximum der Häufigkeitsverteilung der Testsequenz zeigt sich für die Stopp-Anteile $M = 0$ Nm. Ebenso zeigt die Verteilung der Originaldaten eine Anhäufung bei $M = 35$ Nm. Die transformierten relativen Rangfolgen $r(M)$ stellen eine Gleichverteilung der Datenpunkte in dem zugehörigen Histogramm dar.

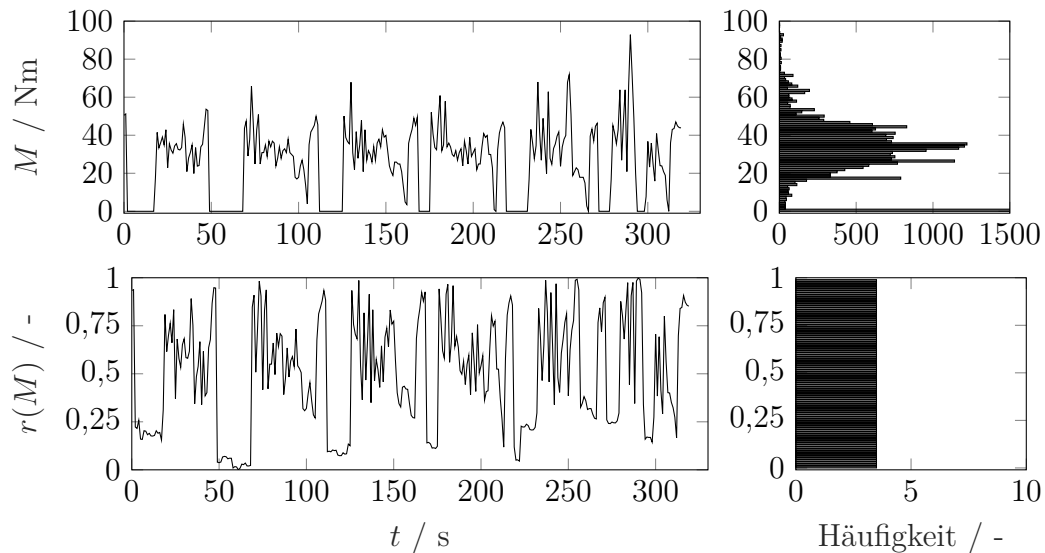


Abbildung 3.2 Schwankungen des Drehmoments M einer RDE-Sequenz mit der Verteilung der Häufigkeit des Auftretens für die Originaldaten M und die relativen Rangfolgen $r(M)$

Die Charakteristik durch die Transformation $t : u(k) \rightarrow R(k)$ (Gl. B.15 im Anhang) ist weitgehend unverändert. Die Frequenzspektren der Zeitreihe von $M(t)$ und der relativen Rangfolge $r(M(t))$ sind zur Verdeutlichung in Abb. 3.3 dargestellt. In der *Fast-Fourier-Transformation* (FFT) ist zu erkennen, dass im Frequenzband von $f = [50, 90]$ Hz keine Verschiebung der Spektren durch die Transformation vorliegt und nur die Amplituden verschoben sind. Durch die Rangtransformation sind die Spitzen im Spektrum bei $r(M(t))$ deutlicher zu erkennen. Dies ist entscheidend bei der Schätzung der Transinformation mit Messdaten, die in einem vorherigen Schritt durch die Rangtransformation gleichverteilt werden. Durch den Schritt ist die relative Häufigkeitsverteilung des Datensatzes gleichverteilt im Bereich $[0, 1]$. Die Charakteristik der Frequenzanteile bleibt im Vergleich den Anteilen der Ursprungsdaten ähnlich.

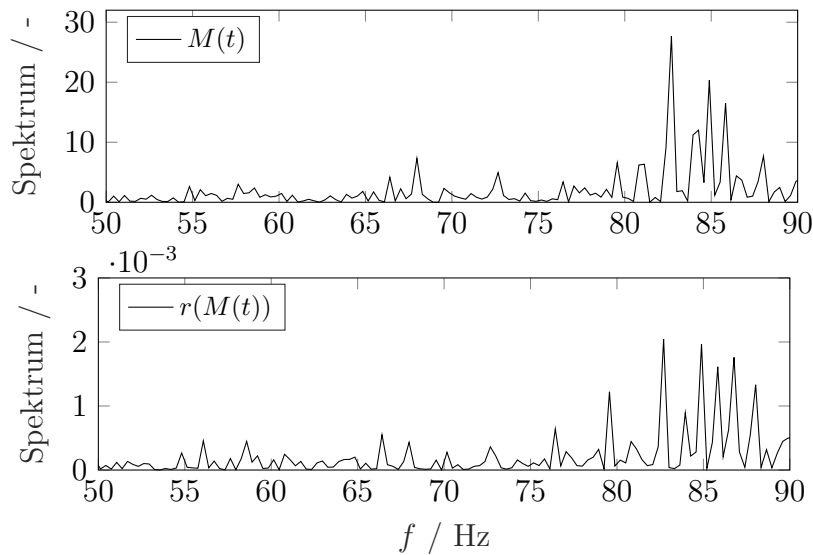


Abbildung 3.3 Frequenzanteile der FFT-Analyse des Drehmomentenverlaufs im Originalsignal $M(t)$ und der transformierten relativen Rangfolgen $r(M(t))$ im Frequenzband $f = [50, 90]$ Hz

Vergleich der Korrelation generischer Daten

Zum Vergleich der Methoden der Datenanalysen werden generische Daten mittels zweier Polynomfunktionen erzeugt und diese anschließend verrauscht. Der Zusammenhang der generischen Daten wird mit den statistischen Maßen nach *Pearson*, nach *Spearman* sowie mit der verallgemeinerten Transinformation bewertet. Die konstruierten Datensätze sind in Abb. 3.4 dargestellt. Der erste Fall in Abb. 3.4a liefert einen nichtlinearen Verlauf, der mit einer quadratischen Funktion approximiert werden kann. Fall zwei zeigt einen stark nichtlinearen Anwendungsfall in Abb. 3.4b.

Die Ergebnisse der statistischen Maße sind in Tab. 3.2 zusammengefasst. Der Fall (a) weist einen hohen linearen Zusammenhang auf, da die Beträge der Korrelationskoeffizienten nahe -1 sind. Der Wert der Transinformation ist unwesentlich kleiner als die Beträge der Ergebnisse der Korrelationskoeffizienten und bestätigt den statistischen Zusammenhang. Die Datenverteilung in Abb. 3.4 ist folglich nach den Korrelationskoeffizienten mit einer linearen Funktion approximierbar.

Im Fall (b) kann hingegen keine lineare Beziehung der Datenmengen ermittelt werden. Die Werte der Korrelationsmaße sind nahe null. Das Ergebnis der Schätzung der Transinformation liegt über den Korrelationskoeffizienten. Durch den vergleichsweise hohen Wert der verallgemeinerten Transinformation \hat{I}_2 lässt sich auf eine statistische Abhängigkeit mit nichtlinearem Verlauf zwischen u und y schließen. Somit bietet die informationstheoretische Betrachtung eine geeignete Methode, um auch nichtlineare Zusammenhänge zu identifizieren.

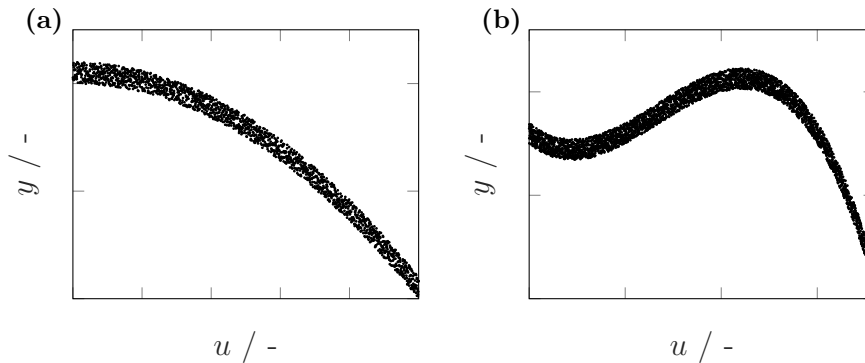


Abbildung 3.4 Zufallsverteilung von Datensätzen zum statistischen Vergleich von Korrelationsanalyse und Transinformation:

- (a) mit einer quadratischen Abhängigkeit $y = -0,2 u^2$
- (b) mit einer kubischen $y = -0,2 u^3 - 0,2 u^2 + 0,05 u$

Tabelle 3.2 Statistische Maße zur Beurteilung der Datenmengen für Fall (a) und (b)

Methoden der Datenanalyse		Fall (a)	Fall (b)
<i>Pearson</i> -Korrelationskoeffizient	r_P	-0,96	-0,17
<i>Spearman</i> -Korrelationskoeffizient	ρ_{SP}	-0,99	0,06
Verallgemeinerte Transinformation	\hat{I}_2	0,83	0,63

Speziell im Versagensfall der Korrelationskoeffizienten bietet die Transinformation ein geeignetes Instrument, Datensätze auf statistische nichtlineare Zusammenhänge zu überprüfen. Verschwindet die Transinformation und liefert $\hat{I}_2 = 0$, sind die Datensätze u und y statistisch unabhängig. Zudem zeigt ein Anwendungsbeispiel in [38, S. 44], dass die Transinformation in der Lage ist, einen statistischen Zusammenhang auch in Datenmengen aufzudecken, die sich auf verschiedene Datenpunkt-Cluster aufteilen.

Datenanalyse von Messdaten

Für eine Gegenüberstellung der Methoden der Datenanalyse werden die Messdaten eines PKW-Dieselmotor-Prüfstandversuchs herangezogen. Dabei sind die Versuchsdaten mit einem statistischen Versuchsplan erstellt worden. An verschiedenen Betriebspunkten im Drehzahl-Last-Kennfeld werden die Kalibrierparameter variiert und die Systemausgänge bzw. Motorgrößen gemessen. Die Korrelationsmethoden und die verallgemeinerte Transinformation sollen die Einflüsse der modellbasierten Kalibrierung auf eine Ausgangsgröße quantifizieren. Die Variationsparameter an einem Drehzahl-Last-Betriebspunkt sind in Tab. 3.3 zusammengefasst.

Tabelle 3.3 Kalibrierparameter einer modellbasierten Kalibrieraufgabe in der Grundmotorenkalibrierung. Die Abkürzung KW steht für Kurbelwinkel

#	Kalibrierparameter	Akronym	Einheit
1	Einspritzzeitpunkt der Haupteinspritzung (EZMI)	ϕ_{MI}	°KW
2	Raildruck	p_{Rail}	bar
3	Rate der Abgasrückführung	ψ_{AGR}	%
4	Einspritzmenge der Nacheinspritzung	q_{post}	mg Hub ⁻¹
5	Abgasladedruck	p_{ALD}	mbar

Die Korrelationsanalyse bezieht sich im linearen Anwendungsfall auf die Ausgangsgröße NO_x-Massenstrom \dot{m}_{NO_x} . Die Ergebnisse der Analyse sind in Abb. 3.5a dargestellt. Die Beziehungen von dem Einspritzzeitpunkt der Haupteinspritzung (EZMI) ϕ_{MI} und dem Abgasladedruck p_{ALD} zu der Ausgangsgröße \dot{m}_{NO_x} weisen die höchste lineare Korrelation mit Werten über 0,6 auf. Die Korrelationskoeffizienten nach *Pearson* und nach *Spearman* sind für sämtliche Größen auf ähnlichem Wertenniveau. Daraus kann auf einen linearen monotonen Zusammenhang für die zwei höchsten Werte geschlossen werden. Die Transinformation gewichtet die Größe des Abgasladedrucks p_{ALD} am höchsten, wertet allerdings die Einflüsse von Raildruck p_{Rail} , Rate der Abgasrückführung ψ_{AGR} und Einspritzmenge der Nacheinspritzung q_{post} höher als die Korrelationskoeffizienten.

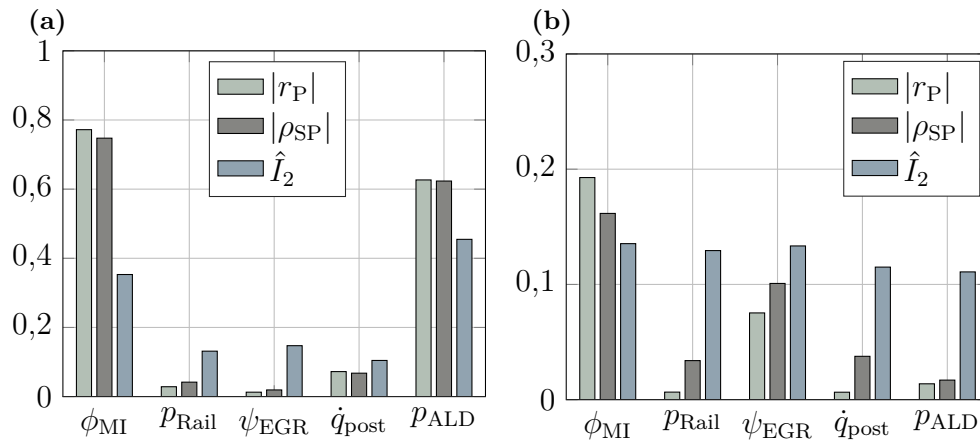


Abbildung 3.5 Vergleich der Korrelationsanalyse der 5 Kalibrierparameter auf die Ausgangsgrößen NO_x-Massenstrom \dot{m}_{NO_x} (a) und effektiver Mitteldruck p_{eff} (b)

Die Korrelationskoeffizienten der fünf Kalibrierparameter zum effektiven Mitteldruck p_{eff} ergeben Werte nichtlinearer Art $< 0,2$. Die Werte der Transinformation sind, ausgenommen der Eingangsparameter EZMI ϕ_{MI} , höher als die der Korrelationskoeffizienten. Den erhöhten nichtlinearen Charakter der restlichen Größen

beweist auch $r_P < \rho_{SP}$. In beiden Anwendungen erweist sich der Parameter ϕ_{MI} als entscheidender Faktor für die beiden Ausgangsgrößen \dot{m}_{NO_x} und den effektiven Mitteldruck p_{eff} . Das Anwendungsbeispiel zeigt die Vorteile der verallgemeinerten Transinformation, welche nichtlineare Zusammenhänge berücksichtigt.

Informationstheoretische Wahl der Abtastzeit

Die Kalibrierung von Motorsteuergeräten verwendet oftmals Messgrößen im Steuergerät für Messungen am Antriebsstrang. Ein Kalibrierwerkzeug erlaubt das Schreiben in und das Lesen aus dem Steuergerät bei einer bestimmten Abtastrate, die frei wählbar ist [59, S. 154]. Die Datenerfassung unterzieht die Messsignale einer Tiefpass-Filterung, wobei diese Anti-Aliasing-Filterung der Messsignale vor der eigentlichen Analog-Digital-Wandlung erforderlich ist [12, S. 503]. Um das Auftreten von Aliasing-Fehlern zu verhindern, bzw. wenn keine Informationen über die Systemdynamiken bekannt sind, wird bei den Messungen über ein Motorsteuergerät mit einer maximal möglichen Abtastrate gemessen. Sollte für eine Datenkomprimierung im Zeitbereich eine geringere Abtastrate benötigt werden, werden die Messsignale nachträglich durch eine Tiefpass-Filterung und anschließend mittels *Downsampling* reduziert.

Speziell in Anwendungen zur Systemidentifikation der in Kap. 2.2 vorgestellten dynamischen Modelle ist die Wahl der Abtastung essenziell. Die Wahl der Abtastzeit spielt eine Rolle für den Frequenzbereich des Modells, welches das Verhalten des Systems beschreiben soll. Gründe hierfür sind zum einen die steigende Anzahl an Parametern bei der Wahl einer zu kleinen Abtastzeit und damit verbunden oftmals numerische Probleme in der Parameterschätzung. Darüber hinaus kann im Extremfall $y(k+1) \approx y(k)$ bei einer Überabtastung kein dynamisches Verhalten zwischen den Ein- und Ausgängen identifiziert werden, wenn die Abtastfrequenz gegenüber der Systemdynamik sehr hoch gewählt ist [85, S. 691, 27, S. 19]. Die Signalwerte der aufeinander folgenden Abtastzeitpunkte unterscheiden sich bei schneller Abtastung kaum [12, S. 503]. Gerade im Modelltraining bei (N)ARX-Modellen werden hohe Frequenzen betont, wodurch diese Modellansätze im Vergleich zu (N)OE-Modellen sensibler auf die Wahl der Abtastung reagieren [27, S. 16, 85, S. 460].

Die Betrachtung zeitdiskreter Signale erfasst Signalwerte zu äquidistanten Zeitpunkten, wobei Signalwerte zwischen den Zeitpunkten verloren gehen [58, S. 163]. Die Abtastzeit T_s bezeichnet die Zeitdauer zwischen zwei Abtastpunkten, bzw. liefert den Kehrwert, die Abtastfrequenz f_s . Nach dem Abtasttheorem von Shannon in Gl. 3.13 gilt für die Abtastfrequenz f_s , dass diese größer als die Hälfte der maximal enthaltenen Frequenz f_{max} sein muss. Die sog. Nyquist-Frequenz $f_s = 2f_{max}$ entspricht der höchsten Frequenz des Signalspektrums [58]. Sind die Voraussetzungen des Abtasttheorems für die Signalabtastung nicht erfüllt, kommt es zu spektraler Überlappung der Spektren des originalen und des abgetasteten Signals und es entsteht Aliasing [59, S. 152].

Als Richtwert für die Festlegung der Abtastfrequenz f_s gilt, dass diese das fünf- bis zehnfache der maximalen Frequenz betragen soll [59, S. 152, 85, S. 512].

$$f_s = \frac{1}{T_s} \geq 2f_{\max} \quad (3.13)$$

Die Datenerfassung tastet die Signale hochfrequent ab. Vor der Modellbildung mit entsprechender Systemidentifikation werden die Signale durch eine digitale Tiefpass-Filterung und abschließender Reduzierung der Abtastfrequenz vor-verarbeitet. Die Erhöhung der Abtastzeit erfolgt auf ein ganzes Vielfaches der Original-Abtastung. Die digitale Tiefpass-Filterung mit einem *Finite Impuls Response* (FIR-) Filter erfolgt mit der Software *Matlab*.

Eine Abtastfrequenz sollte nicht zu hoch gewählt werden, da Zustandswechsel Zeit benötigen. Eine zu kleine Wahl der Abtastfrequenz kann hingegen Zustände verpassen und dabei Information verlieren [89, S. 96]. Dies soll im Folgenden für zwei RDE-Datensätze an einem Beispiel mittels der verallgemeinerten Transinformation und der Entropien erläutert werden: Die Entropien der beiden Datensätze sind das Maß für den mittleren Informationsgehalt. Die Verbundentropie aus Gl. 3.6 berechnet sich mit der Wahrscheinlichkeit des Auftretens der paarweisen Ereignisse u_i und y_j und umfasst den gesamten Informationsfluss $H(\mathbf{U}, \mathbf{Y})$. Dabei wird bei den Messdaten von statistisch abhängigen Ereignissen ausgegangen und der Einfluss der Datenkomprimierung, des *downsampling*, mit der Gesamtinformation anhand zweier Signalquellen verglichen. Zu prüfen ist die Stärke des Einflusses bei der Komprimierung mit den Informationsmaßen, die in Abb. 3.6 für die Signalquellen einer Lastgröße, das Drehmoment M und den NO_x -Massenstrom \dot{m}_{NO_x} geschätzt werden. Die Datengrundlage basiert auf zwei RDE-Ersatzzyklen mit den Indizes 1 und 2.

Zur Verdeutlichung sind die vier Diagramme in Abb. 3.6 auf der Achse der Abtastzeit T_s logarithmisch dargestellt. Die Signalquellen $Y = \dot{m}_{\text{NO}_x}$ werden für die Berechnung auf eine Gleichverteilung rangtransformiert. Jeweils für eine steigende Abtastzeit von $T_s = 0,02 \text{ s} - T_s = 5 \text{ s}$ werden die Informationsmaße und Entropien berechnet. In Abb. 3.6a ist die Entropie der zur Gleichverteilung rangtransformierten Messdaten \dot{m}_{NO_x} gezeigt. Die Werte sind auf konstantem Niveau für eine steigende Abtastzeit bis zu einer Abtastung von $T_s = 3 \text{ s}$. Ab diesem Grenzwert ist der Prozess des Downsamplings für dieses Beispiel nicht mehr verlustfrei.

Die Schätzungen der nichtnormierten verallgemeinerten Transinformation \hat{I}_2 nach Gl. 3.12 sind in Abb. 3.6b dargestellt. Die Transinformation steigt kontinuierlich mit sinkender Abtastung für beide RDE-Datensätze an. Das Level des RDE-Ersatztest mit dem Index 1 liegt aufgrund der höheren Dynamik des zweiten Datensatzes über dem mit Index 2. Mit steigender Dynamik sinkt der statistische Zusammenhang der beiden Datenquellen M und \dot{m}_{NO_x} . Die Transinformation I_2 beider Sequenzen nähert sich dem Wertniveau der Entropie $H_2(\dot{m}_{\text{NO}_x})$ an. Aus den Eigenschaften der Transinformation in Tab. 3.1 folgt, dass die Wahrscheinlichkeitsverteilung von \dot{m}_{NO_x} eine Funktion der Verteilung von M wird. Dies zeigt sich für eine steigende Abtastzeit T_s ab $0,2 \text{ s} - 0,5 \text{ s}$.

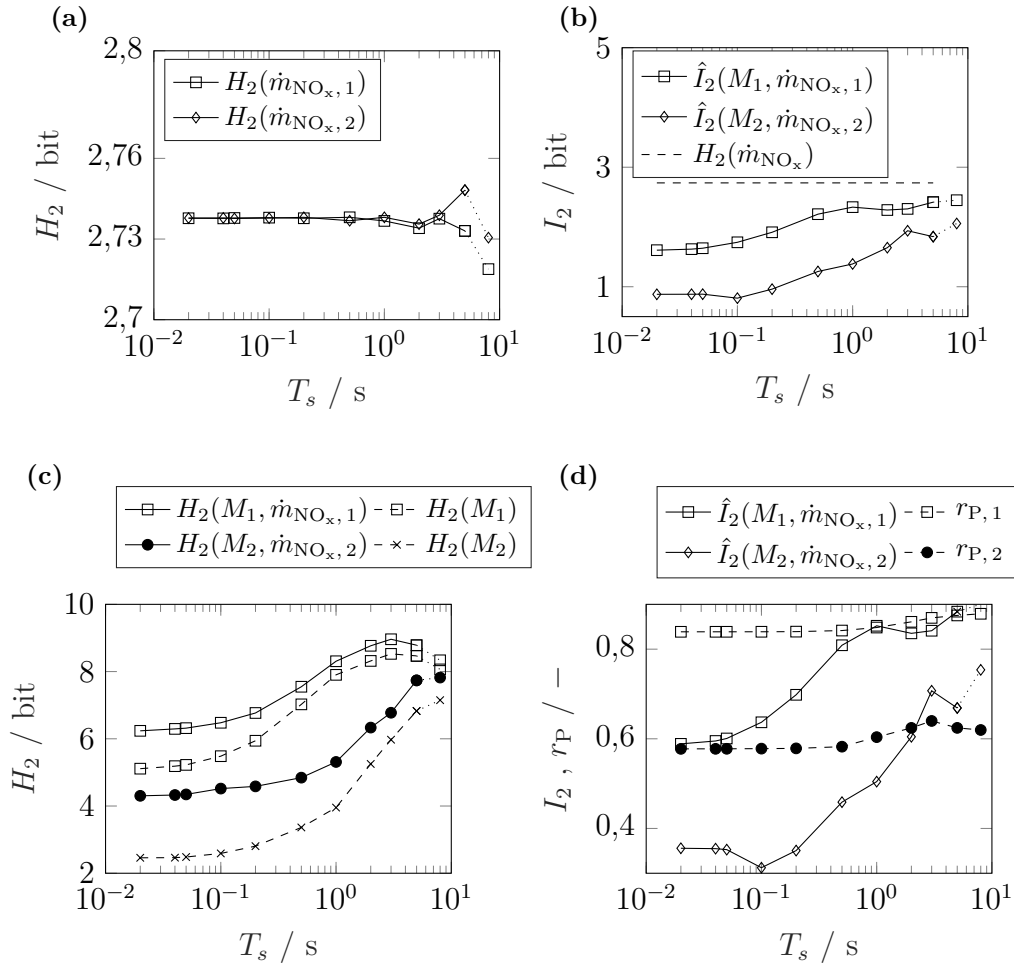


Abbildung 3.6 Einfluss der Abtastzeit T_s auf die Informationsmaße zweier verschiedener RDE-Ersatzzyklen für die beiden Signale Drehmoment M und NO_x -Massenstrom \dot{m}_{NO_x}

In Abb. 3.6c sind die Entropie $H_2(M)$ sowie die Verbundentropie $H_2(M, \dot{m}_{\text{NO}_x})$ für beide RDE-Datensätze dargestellt. Mit der sinkenden Abtastung steigt die Verbundentropie $H_2(M, \dot{m}_{\text{NO}_x})$ bis $T_s = 3$ s an; für ein weiter steigendes T_s nimmt sie wieder ab. Gleiches gilt für die Entropie $H_2(M)$. Auffällig ist zudem, dass die Abstände zwischen der Entropie sowie der Verbundentropie für die Fallbeispiele geringer werden und die Entropie $H_2(M)$ gegen den Wert der Verbundentropie $H_2(M, \dot{m}_{\text{NO}_x})$ strebt. Das liegt an der Schätzung der Transinformation \hat{I}_2 (Gl. 3.12) und den Eigenschaften der Transinformation (Tab. 3.1). Bei $\hat{I}_2(\mathbf{U}; \mathbf{Y}) \rightarrow H_2(\mathbf{Y})$ folgt $H_2(\mathbf{U}) \rightarrow H_2(\mathbf{U}, \mathbf{Y})$. Durch die Datenkomprimierung des Downsamplings sinkt die Anzahl an Zeichen des Senders $H_2(\mathbf{U})$. Im Informationskanal Abb. 3.1 steigt damit die Unsicherheit an der Quelle für den mittleren Informationsgehalt $H_2(\mathbf{U})$. Mit steigender Unsicherheit (Entropie) sinkt der Informationsgehalt über die gesendeten Zeichen.

Die normierte verallgemeinerte Transinformation \hat{I}_2 in Abb. 3.6d steigt mit T_s gegen den Wert eins und die statistische Abhängigkeit der Datenquellen wird maximal. Wird der nichtlineare statistische Zusammenhang der Transinformation \hat{I}_2 mit dem Maß des linearen Zusammenhangs nach *Pearson* r_P verglichen, fällt auf, dass der lineare Zusammenhang trotz der Änderung der Abtastfrequenz nahezu konstant bleibt. Der Korrelationskoeffizient erfasst die Veränderungen der Wahrscheinlichkeitsverteilung durch die Datenkomprimierung nicht. Aus der Betrachtung der Informationsmaße folgt eine geeignete Abtastung T_s für die Datensätze der beiden RDE-Ersatztests zwischen 0,02 s und 0,2 s, da in diesem Bereich die Ähnlichkeit der Informationsmaße nicht gravierend durch die Datenkomprimierung beeinträchtigt sind. Das Ergebnis deckt sich mit der in Kap. 3.4.5 vorgestellten Analyse der Messtechnik des NO_x -Sensors, der in diesem Beispiel für die Aufnahme des Emissionsmassenstroms herangezogen wird. Eine physikalisch sinnvolle Wahl der Abtastung für beispielsweise die Systemidentifikation einer dynamischen Modellstruktur kann folglich mittels einer informationstheoretischen Anschauung überprüft werden.

3.2 Methoden der Sensitivitätsanalyse

Ziel der Sensitivitätsanalyse ist es, die Einflüsse und Beziehungen der Eingangs- und Ausgangsgrößen eines beliebigen Modellansatzes sowie deren Wechselbeziehung zu quantifizieren [102]. Die Korrelationsanalyse vergleicht die statistischen Beziehungen verschiedener Datensätze miteinander. Die Sensitivitätsanalyse untersucht den Einfluss der Varianzen in den Modelleingängen auf die Varianz am Modellausgang [102, S. 45]. Anstelle der Rohdaten wird die Wirkung der Modelleingänge auf die Ausgänge betrachtet. Zu den weit verbreiteten Methoden gehören das *Factor-Screening* sowie die lokale und die globale Sensitivitätsanalyse [60, S. 216, 113, 102, S. 161, 31]. *Factor-Screening* misst die qualitativen Einflüsse von Eingangs- auf Antwortkanäle und kann dadurch signifikante von nichtsignifikanten Parametern unterscheiden. Die Betrachtung des Einflusses durch die Variation eines einzelnen Eingangsparameters aus dem Versuchsraum bei gleichzeitigem Festhalten der anderen Einflussgrößen wird als lokal bezeichnet [113, S. 11, 47, S. 10]. Lokale, eindimensionale Bewertungsmaße werden mit Differenzen- und Differentialquotienten gebildet. Die globale Sensitivitätsanalyse dagegen betrachtet den gesamten Definitionsbereich des Eingangsparametergebiets.

Aus der Gruppe der globalen Ansätze werden die **varianzbasierten Methoden** nach Sobol und der *Fourier amplitude sensitivity test* (FAST) vorgestellt. Eine visualisierte globale Analyse von Systemzusammenhängen zwischen Ein- und Ausgängen bieten sog. **visuelle Erklärbarkeitsansätze**. In Kap. 2.1 wird die empirische Modellbildung mit Polynommodellen und GPM vorgestellt, bei denen die Parameter der jeweiligen Schätzverfahren ebenfalls als Sensitivitätsmaße eingesetzt werden können. Deren Güte kann im Falle von Nichtlinearitäten bei nicht-monotonen Modellen eingeschränkt sein [102, S. 43].

3.2.1 Varianzbasierte Methoden

Die Methoden der varianzbasierten Sensitivitätsanalyse sind sowohl auf lineare als auch nichtlineare Modelle anwendbar. In der Analyse werden sowohl die Verteilungsfunktionen der Eingangsgrößen als auch die Variation aller Eingangsgrößen berücksichtigt. Die varianzbasierte Betrachtungsweise baut auf einer Stichprobengenerierung auf, vorwiegend mittels Monte-Carlo-Methoden. Die Schritte des generellen Ablaufs der auf Stichproben basierenden Sensitivitätsanalyse gliedern sich wie folgt [113, 102]:

1. Definition der Eingangsraumgrenzen und Stichprobengenerierung für den gesamten Eingangsraumbereich
2. Berechnung der Modellausgänge für die jeweiligen Stichproben des Modelleingangs und Analyse der Modellausgangsvarianzen
3. Sensitivitätsanalyse: Ermittlung des Einflusses der jeweiligen Eingangsvariationen auf die Ausgangsvarianzen

Die varianzbasierte Sensitivitätsanalyse bestimmt den Anteil der Varianz des Ausgangs, der durch einen Eingang verursacht wird. Folglich ist ein Vergleich der Signifikanz der verschiedenen Eingänge auf die Ausgangsvariablen möglich [115, S. 416]. Zur Bestimmung der Sensitivitätsmaße werden bedingte Varianzen berechnet, die sich auf den von einer Eingangsgröße erzeugten Anteil an der Varianz der Ausgangsgröße beziehen [113, S. 28]. Die varianzbasierte Sensitivitätsanalyse folgt der Annahme, dass ein wahrer Wert \tilde{u}_i einer Eingangsgröße u_i bekannt ist, und bestimmt eine bedingte Varianz $\sigma_{\mathbb{E}\{Y|\tilde{u}_i\}}^2$, die den wahren Wert des Eingangsparameters beschreibt. Da der genaue Eingangswert nicht bekannt ist, wird der Erwartungswert $\mathbb{E}\{Y|u_i\}$ über dem gesamten Eingangsvariationsbereich von u_i zur Bestimmung des globalen Sensitivitätsmaßes verwendet. Darauf aufbauend wird im weiteren Verlauf des Kapitels entsprechend der gebräuchlichen Literatur vereinfacht die bedingte Varianz $\sigma_{\mathbb{E}\{Y|u_i\}}^2$ mit $\sigma_{Y|u_i}^2$ ersetzt [113, 102]. Die Sensitivitätsmaße oder -indizes werden als Quotient des Anteils der Varianz einer Eingangsgröße $\sigma_{Y|u_i}^2$ auf die Varianz der Ausgangsgröße σ_Y^2 bestimmt:

$$S_i = \frac{\sigma_{Y|u_i}^2}{\sigma_Y^2} \quad (3.14)$$

Der Quotient S_i der Gl. 3.14 wird teilweise als Korrelationsverhältnis teilweise als Sensitivitätsindex der ersten Ordnung bezeichnet und erfasst die Sensitivität der Haupteffekte eines additiven Modells vollständig. Additive Modelle beschreiben einen linearen Zusammenhang der Eingangsgrößen mit einem Modell erster Ordnung [59, S. 46]. Bei unkorrelierten Eingangsparametern gilt für die Summe der jeweiligen Sensitivitätsindizes der Haupteffekte $\sum_{i=1}^{n_p} S_i = 1$, da jedes S_i einen direkten quantitativen Beitrag des Eingangs u_i auf die Ausgangsvarianz σ_Y^2 beschreibt.

Für nichtlineare Zusammenhänge oder nichtadditive Modelle müssen diese Beziehungen nicht gültig sein. Der Quotient kann modellunabhängig genutzt werden, um beispielsweise die Additivität von Modellen zu identifizieren. Für die globale Sensitivitätsanalyse bei nichtadditiven Modellen werden die Interaktionseffekte der Eingänge untereinander berücksichtigt. Die Effekte höherer Ordnung reichen von der zweiten bis zur n_p -ten Ordnung. Die Sensitivitätsanalyse nach Sobol bildet den nichtlinearen funktionalen Zusammenhang zwischen den Eingängen \mathbf{u} und dem Ausgang y mittels separater Funktionen in Summanden steigender Ordnungen ab [117]. Jeder Term gibt eine mögliche Faktorkombination u_1, \dots, u_{n_p} wieder, die wiederum Sensitivitätsmaße höherer Ordnungen darstellt (siehe die Herleitung Kap. B.5 im Anhang).

Die Berechnung sämtlicher Sensitivitätsindizes höherer Ordnungen besonders für komplexe Modelle ist mit einem enormen Rechenaufwand verbunden. Abhilfe schafft die Betrachtung des totalen Sensitivitätsindex S_{T_i} , bei dem alle Eingangsgrößen bis auf u_i festgehalten werden. [102] berechnet die totalen Effekte der Eingangsgröße u_i mit Hilfe der bedingten Varianz

$$\sigma_{Y|u_1, u_2, \dots, u_{i-1}, u_{i+1}, \dots, u_{n_p}}^2 = \sigma_{Y|u_{\sim i}}^2. \quad (3.15)$$

Mit der bedingten Varianz $\sigma_{Y|u_{\sim i}}^2$, bei der alle Eingangsgrößen bis auf u_i festgehalten werden, kann der modellunabhängige totale Sensitivitätsindex S_{T_i} (Gl. 3.17) berechnet werden. Hier sind die Formeln der Sensitivitätsanalyse für Haupt- und Totaleffekte gegenübergestellt:

$$\text{Haupteffekt:} \quad S_{H_i} = S_i = \frac{\sigma_{Y|u_i}^2}{\sigma_Y^2} \quad (3.16)$$

$$\text{Totaleffekt:} \quad S_{T_i} = \frac{\mathbb{E}\{\sigma_{Y|u_{\sim i}}^2\}}{\sigma_Y^2} = 1 - \frac{\sigma_{Y|u_{\sim i}}^2}{\sigma_Y^2} \quad (3.17)$$

Für $S_{T_i} = 0$ beeinflusst der Faktor u_i den Modellausgang nicht. Für $S_{T_i} \gtrsim 0$ hat der Faktor auf den Ausgang keinen wesentlichen Einfluss [102, S. 163]. Die Differenz zwischen den Totaleffekten und den Haupteffekten bestimmt die Interaktionseffekte. Eigenschaften der Indizes für additive und nichtadditive Modelle sind in Tab. 3.4 zusammengefasst.

Tabelle 3.4 Eigenschaften der Sensitivitätsindizes [113, S. 33]

Modellart	Haupteffekt	Totaleffekt
Additiv	$\sum_{i=1}^n S_i = 1$	$\sum_{i=1}^n S_{T_i} = 1$
Nichtadditiv	$\sum_{i=1}^n S_i \leq 1$	$\sum_{i=1}^n S_{T_i} > 1$

Die Schätzung der bedingten Varianzen $\sigma_{Y|u_i}^2$ und $\sigma_{Y|u_{\sim i}}^2$ erfolgt mit der Monte-Carlo-Methode (Kap. B.5.1) oder mit dem FAST (Kap. B.5.2) [102].

Bei der Monte-Carlo-Methode werden gleichverteilte Zufallsstichproben im Eingangsraum-Hyperkubus des Versuchsraums erzeugt. Aus den Modellauswertungen für den Stichprobenumfang werden die Ausgangsvarianzen sowie die bedingten Varianzen geschätzt [102, S. 164ff]. Die FAST-Methode erzeugt mittels Anregungsfunktionen für jeden Eingang einen sinusförmigen Stichprobenumfang, wobei ein jeder Eingang eine unterschiedliche eigene, vordefinierte Frequenz zugeordnet bekommt. Die Frequenzen sind dabei linear unabhängige Vielfache voneinander und werden im Allgemeinen nach einer Tabelle ausgewählt [25]. Der Stichprobenumfang der FAST-Methode regt die Modellausgänge an. Der Ausgang wiederum wird mit einer Fourier-Reihenentwicklung beschrieben, und aus den Fourier-Koeffizienten werden die Schätzwerte der Varianz des Modellausgangs und die bedingte Varianz $\sigma_{Y|u_i}^2$ des Eingangs u_i berechnet [113, S. 36, 102, S. 167, 25].

3.2.2 Visuelle Erklärbarkeitsansätze

Die grafische Auswertung von *Black box*-Modellprädiktionen mittels sogenannten Erklärbarkeitsansätzen ist im maschinellen Lernen weit verbreitet. Erklärbarkeitsansätze helfen dabei, Einflüsse einzelner Eingänge auf den Modellausgang zu verstehen. Diese Einflüsse können qualitativ per Visualisierung mit *Partial dependence plots* (PDP) beschreiben werden. PDPs visualisieren die Vorhersagen eines Modells abhängig von einer oder mehreren selektierten Variablen sowie die Änderungsrate der Modellvorhersage, während über alle anderen unabhängigen Eingangsvariablen integriert wird [37]. Ein PDP kann dabei feststellen, ob Zusammenhänge linear, monoton oder nichtlinear sind. Die partiellen Abhängigkeitsdiagramme sind aufgrund der Limits einer grafischen Visualisierung auf ein- bis zweidimensionale Beziehungen limitiert [37, 109, 7, S. 82]. Die *Partial dependence* (PD) ist definiert als [44, 109]

$$y_{\text{PD}}(\mathbf{u}_s) = \mathbb{E}\{\hat{y}(\mathbf{u}_s, \mathbf{U}_{\sim s})\}. \quad (3.18)$$

Es sei \mathbf{u}_s der Eingang, an dem der Einfluss auf die Modellprognose untersucht wird, und $\mathbf{U}_{\sim s} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_{s-1} \ \mathbf{u}_{s+1} \ \dots \ \mathbf{u}_{n_p}]^\top$ der Zusammenschluss der restlichen Eingänge des Modells, über die integriert wird. Die Matrix $\mathbf{U}_{\sim s}$ und der Vektor \mathbf{u}_s bilden den gesamten Versuchsraum $\mathbf{U} = [\mathbf{u}_s, \mathbf{U}_{\sim s}]$ ab. Eine grafische Auswertung der Modellausgänge über der Datenmenge wird mit PDP bezeichnet. Jede Eingangsdimension u_s besitzt eine eigene PD (Gl. 3.18), die einen durchschnittlichen Ausgabewert über die gesamte Wertemenge des Versuchsraums liefert. Alle verbleibenden Eingangsvariablen des Modells sind in $\mathbf{u}_{\sim s}$ repräsentiert. Da die marginale Verteilung der Modellvorhersage von $\mathbf{u}_{\sim s}$ nicht bekannt ist, berechnet sich die PD über die Schätzfunktion durch eine Monte-Carlo-Integration mit [44, 109]

$$\hat{y}_{\text{PD}}(\mathbf{u}_s) = \sum_{i=1}^N \hat{y}_s(\mathbf{u}_s) = \frac{1}{N} \sum_{i=1}^N \hat{y}(\mathbf{u}_s, \mathbf{u}_{\sim s}^{(i)}). \quad (3.19)$$

Die Approximation am Punkt $\mathbf{u}_{\sim s}^{(i)}$ entspricht dem arithmetischen Durchschnitt des i -ten Zeilenelements der Prädiktionsmatrix und der tatsächlichen Ausprägung der nicht betrachteten Variablen des Datensatzes. Formal berechnet der PDP die durchschnittliche Modellvorhersage über alle möglichen Ausprägungen in \mathbf{u}_s und ist ein globaler Ansatz für funktionelle Zusammenhänge [44, 109].

Da der PDP keine Interaktionseffekte berücksichtigt, wird die Erweiterung *Individual conditional expectation* (ICE) eingeführt [44]. Anstelle der durchschnittlichen Änderung der Vorhersage berücksichtigt die Darstellungsform der ICE die Änderung der Vorhersage bei der Manipulation der Merkmalausprägung jeder Instanz. Ein ICE-Plot ist folglich eine Aufschlüsselung des PDPs für alle möglichen Ausprägungen in \mathbf{u}_s und resultiert in einer einzelnen Linie pro Instanz. Generell entspricht ein PDP einer Mittelung sämtlicher Linien des ICE-Plots. Die Berechnung erfolgt ähnlich der Gl. 3.18 mit der Aufteilung des PDPs in N verschiedene Kurven. Der gesamte Eingangsraum wird in $\mathbf{u}_{\sim s}$ konstant gehalten und für jede Instanz i werden die Werte aus einem Stichprobenraster in das Modell gegeben. Die Modellvorhersage gibt Punkte für eine Instanz i aus, die verbunden eine Linie geben. Jede Kurve repräsentiert die Modellvorhersage als Funktion eines Eingangsparameters \mathbf{u}_s bedingt zur Beobachtung $\mathbf{u}_{\sim s}$. Anstelle des durchschnittlichen partiellen Effektes von $\mathbf{u}_{\sim s}$ werden die N geschätzten bedingten Erwartungskurven der Zielvariable berechnet. Die erwartete Vorhersage der Variable nach i -ter Beobachtung in Abhängigkeit von \mathbf{u}_s bedingt auf den beobachteten Vektor $\mathbf{u}_{\sim s}$ ist definiert als [109]

$$\hat{y}_{\text{ICE}}(\mathbf{u}_s) = \mathbb{E}\{\hat{y}(\mathbf{u}_s, \mathbf{u}_{\sim s}^{(i)})\}. \quad (3.20)$$

3.2.3 Anwendungsbeispiele der Sensitivitätsanalyse

Die varianzbasierte Sensitivitätsanalyse wird im Folgenden an einem aus generischen Daten antrainierten Modell erläutert. Zur Schätzung erfolgt die Stichprobengenerierung mittels FAST sowie Monte-Carlo-Methode. Zusätzlich werden die Sensitivitätsindizes mit dem Hyperparameter ℓ eines GPM-Schätzverfahrens verglichen. Der Hyperparameter kann ebenfalls als Sensitivitätsmaß eingesetzt werden. Anschließend wird das Beispiel aus Kap. 3.1.3 aufgegriffen. Anstelle der Rohdaten wird die Modellierung auf die Sensitivität bezüglich der Modellausgänge geprüft. Die erzeugten Modelle werden mittels der visuellen Erklärbarkeitsansätze ICE und PDP visuell ausgewertet, um den Einfluss und die Richtung der Sensitivität zu beschreiben.

Vergleich der Sensitivität an generischen Daten

Die Vor- und Nachteile von varianzbasierten Sensitivitätsanalysen werden nachfolgend anhand eines Modells mit verschiedenen generischen Eingangscharakteristika aufgezeigt. Der Eingangsraum des Beispiels wird im Bereich von 0 bis 1 für jede der vier Eingangsdimensionen begrenzt.

Somit bildet der Raum einen vierdimensionalen Hyperkubus ab. Durch vier verschiedene Funktionen werden generische Zufallsdatenpunkte erzeugt, die zusätzlich um die Funktionswerte gestreut werden. Die Anzahl generisch erzeugter Messdaten ist M . Die vier Modelleingänge werden durch eine Exponentialfunktion in Abhängigkeit von u_1 , durch die Polynomfunktion des Exponenten 10 von u_2 , einen linearen Anteil von u_3 und einer von u_4 abhängigen Polynomfunktion mit dem Exponenten 5 beschrieben. Ein Modellausgang ist entsprechend durch eine Summierung der Funktionen gegeben mit

$$\hat{y}(\mathbf{u}) = \frac{\exp(u_1) - 1}{\exp(1) - 1} + u_2^{10} + u_3 + u_4^5. \quad (3.21)$$

Die einzelnen Eingangsfunktionen der Gl. 3.21 sind in Abhängigkeit von $u_1 - u_4$ in Abb. 3.7a dargestellt. Ausgangsdaten werden in der jeweiligen Eingangsdimension mit der zugehörigen Funktion $y_i = f(u_i)$ aus gleichverteilten Datenpunkten generiert. Anschließend werden die Funktionswerte y_i durch die Addition von normalverteilten Rauschen gestreut. Über die Gl. 3.21 werden mit den gestreuten Datenpunkten Ausgangswerte \hat{y} erzeugt. Das Systemverhalten $\hat{y}(\mathbf{u})$ der erzeugten Eingangs- und Ausgangsdaten wird mit einem GPM approximiert.

Die Genauigkeit und der numerische Aufwand einer Sensitivitätsanalyse hängen vom Stichprobenumfang N der Monte-Carlo- bzw. FAST-Schätzung ab. Die Auswirkung der Größe des Stichprobenumfangs auf den Schätzwert der Gesamtvarianz s_Y^2 des Modellausgangs zeigt die Abb. 3.7b. Ein gleichbleibendes Level der Gesamtvarianz ist für einen Stichprobenumfang von $N > 500$ zu beobachten. Die nachfolgenden Untersuchungen umfassen den Stichprobenumfang $N = 1.000$.

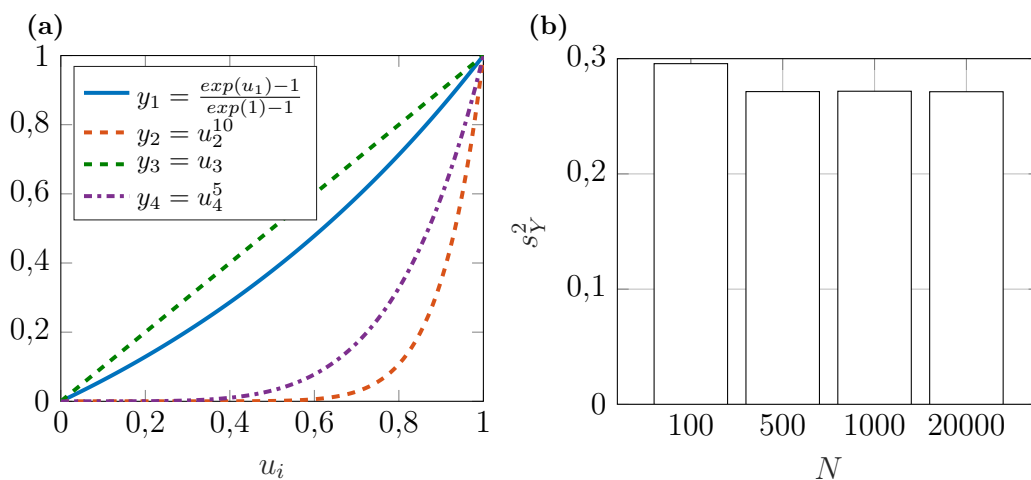


Abbildung 3.7 Überblick der Eingangsfunktionen des Anwendungsfalles (a) und der Einfluss der Stichprobengröße N auf den Schätzwert der Varianz des Gesamtmodells s_Y^2 (b)

Der Längenskalare ℓ_i des GPMs wird als Sensitivitätsmaß eingesetzt und mit den varianzbasierten Indizes verglichen (siehe Kap. 2.1.2). Die Qualität der Modellschätzung hängt stark von der Anzahl an Datenpunkten M ab. Die Größe der Längenskalare ist in Abb. 3.8a gezeigt und variiert mit einer steigenden Anzahl von Datenpunkten M . Die Längenskalare variieren im Bereich von $\ell_1 = 0,9 - \ell_3 = 6,9$. Eine stabile Modellvorhersage gilt für eine Datensatzgröße von $M \geq 200$. Obwohl die Modellqualität und -vorhersage von der Anzahl an Datenpunkten abhängig ist, sind die jeweiligen totalen Sensitivitätsindizes in Abb. 3.8b für jedes GPM nahezu gleich. Die varianzbasierte Sensitivitätsanalyse ist komplett unabhängig von der Modellqualität.

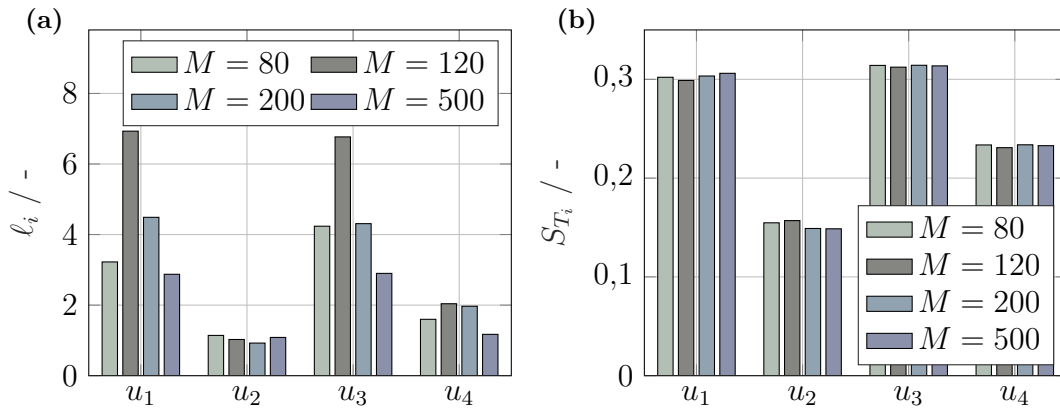


Abbildung 3.8 Der GPM-Hyperparameter – der Längenskalare – variiert stark abhängig von der Datensatzgröße (a), während die Indizes der varianzbasierten Sensitivitätsanalyse S_{T_i} unabhängig von der Anzahl an Messungen und folglich der Modellgüte sind (b)

Die Schätzung der varianzbasierten Sensitivitätsmaße erfolgt mit der GP-Modellierung des Datensatzes $M = 200$ bei einem Stichprobenumfang von $N = 1.000$. In Abb. 3.9a sind die Längenskalare und die Werte der Sensitivitätsindizes für die Eingangsparameter aufgetragen. Ein kleiner Längenskalare ℓ_i ist gleichbedeutend mit einer hohen Sensitivität bezüglich des betrachteten Parameters. Der Hyperparameter wird in einen sich konträr verhaltenden Parameter $\ell_{\text{skaliert}, i}$ umgeformt. Dadurch wird eine Vergleichbarkeit mit den varianzbasierten Indizes erzielt. Die Vergleichsgröße ergibt sich mit

$$\ell_{\text{skaliert}, i} = \frac{1}{n_p} \left(1 - \frac{\ell_i}{\sum_{j=1}^{n_p} \ell_j} \right). \quad (3.22)$$

Die Haupteffekte S_i und $S_{\text{FAST}, i}$ sowie die Gesamteffekte $S_{T, i}$ resultieren unabhängig vom Schätzverfahren in ähnlichen Werten. Folglich ist das Modell additiv. Die mit der Gl. 3.21 erzeugten generischen Daten beinhalten keine Interaktionseffekte der Eingänge und $S_i \approx S_{T, i}$.

Die minimale Differenz der Indizes ist eine Folge des gleichverteilten Stichprobenumfangs der Schätzmatrizen. Die Indizes ordnen die Sensitivität des Eingangs u_i zum Modellausgang in absteigender Reihenfolge: u_3, u_1, u_4 und u_2 , siehe Abb. 3.9a. Der Hyperparameter ℓ_i des geschätzten GPMs ordnet den Modelleinfluss in absteigender Reihenfolge: u_2, u_4, u_3 und u_1 . Der Längenskalar ordnet den Eingangsdaten mit steigender Nichtlinearität niedrigere Werte zu, bzw. höhere skalierte Werte zu. Die Sensitivität des Modells ist damit bezüglich niedriger Werte höher. Die nahe beieinander liegenden Datenpunktkurven des exponentiellen u_1 - und linearen u_3 -Anteils gewichtet der Längenskalar annähernd gleich. Die höchsten Sensitivitäten erzeugen die Eingangspolynomfunktionen. Der Exponent 10 von u_2 senkt das Wertenniveau von ℓ_{skaliert} gegenüber dem Exponenten 5 von u_4 .

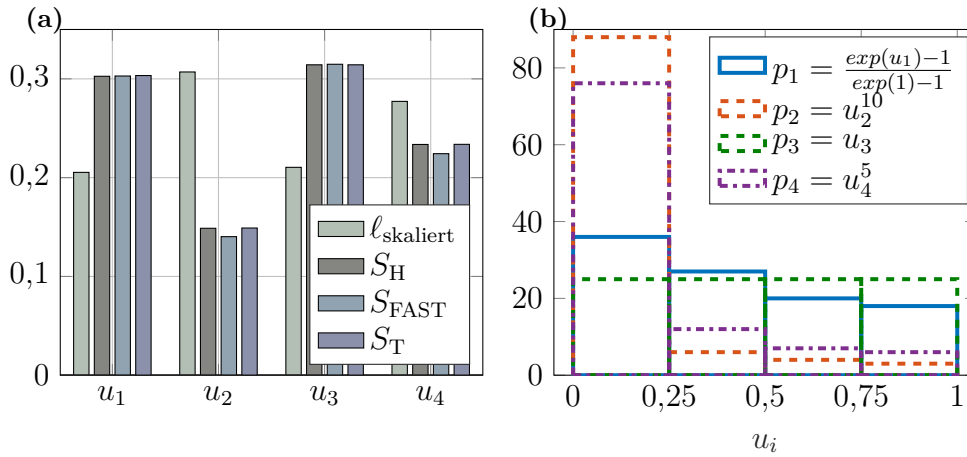


Abbildung 3.9 Ergebnisse der Sensitivitätsanalyse des generischen Anwendungsfalles mit $M = 200$ Punkten (a) und die zugehörigen Wahrscheinlichkeitsverteilungen $p(y_i)$ der verschiedenen Eingangsfunktionen y_i für eine Anzahl von 100 Punkten (b)

Die varianzbasierte Sensitivitätsanalyse generiert zur Monte-Carlo-Schätzung der Indizes einen mehrdimensionalen Stichprobenumfang. Dieser ist gleichverteilt im normierten Hyperkubus des Versuchsraums $[0, 1]$ [117]. Die Schätzung berechnet die Modellvarianz über den gesamten Eingangshyperkubus. Die geometrischen Schwerpunkte der Flächen der Wahrscheinlichkeitsverteilung p_i hängen mit der Schätzung der Varianz und den Sensitivitätsindizes zusammen. Die Wahrscheinlichkeitsdichten $p_i = p(u_i)$ der einzelnen Eingangsvariablen u_i des GPMs sind für einen Stichprobenumfang von $N = 100$ Punkten in Abb. 3.9b dargestellt. Den höchsten Wert des Flächenschwerpunktes liefert der lineare Eingang mit $\bar{p}(u_3) = 0,5$. Die Funktion des Eingangs u_2 mit dem Exponenten 10 visualisiert eine Kurve mit hoher Steilheit in der Eckregion $0,6 < u_i < 1$. Das geometrische Zentrum der Wahrscheinlichkeitsverteilung ist mit $\bar{p}(u_2) = 0,095$ am kleinsten. Der gleichverteilte Stichprobenumfang erzeugt eine von der Eingangscharakteristik abhängige Varianz.

Die Nichtlinearität der Eingangsfunktionen wird in der Monte-Carlo-Simulation des Anwendungsbeispiels in Gl. 3.21 nicht wesentlich bewertet und resultiert in einer niedrigen Sensitivität. Das nichtlineare Eingangsverhalten wird dagegen vom Längenskalarhyperparameter erfasst. Die Sensitivitätsanalyse mit einem GPM berücksichtigt folglich im Gegensatz zu den varianzbasierten Ansätzen auch Nichtlinearitäten.

Sensitivitätsanalyse auf Daten der Grundmotorenkalibrierung

In Kap. 3.1.3 ist eine Korrelationsanalyse auf ein DoE der Grundmotorenkalibrierung beschrieben. Die Datengrundlage dieser Korrelationsanalyse wird nun für die Sensitivitätsanalyse herangezogen. Die Abbildung der Messdaten erfolgt mithilfe einer GP- und einer Polynomstruktur. Die Schätzung der Haupteffekte S_H und Total-effekte S_T für das Polynommodell oder GPM erfolgt mit der Monte-Carlo-Methode. Verglichen werden die Sensitivitätsindizes mit dem skalierten Längenskalar $\ell_{\text{skaliert}, i}$ des GPM. Die Ergebnisse zeigt Abb. 3.10.

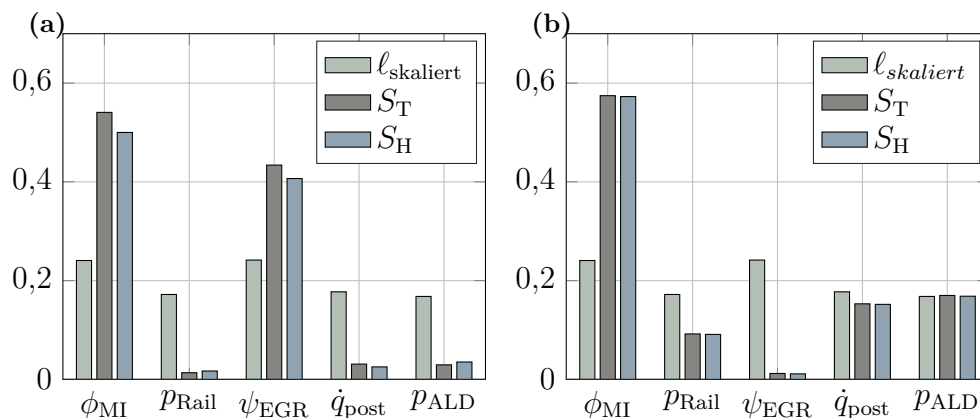


Abbildung 3.10 Ergebnisse der Sensitivitätsanalyse auf die Ausgangsgrößen NO_x -Massenstrom \dot{m}_{NO_x} (a) und effektiver Mitteldruck p_{eff} (b)

Die Haupteinflussfaktoren stehen in einer entgegengesetzten Reihenfolge zu den Resultaten der Korrelationsanalyse. Die beiden Eingangsgrößen EZMI ϕ_{MI} und AGR-Rate ψ_{AGR} kennzeichnen die höchsten Einflüsse des \dot{m}_{NO_x} -Modells (Abb. 3.10a). Die beiden Haupteinflüsse summieren sich auf über 90%. Die Totaleffekte addieren sich auf 97%. Die Differenz der Ergebnisse von Totaleffekt S_T und Haupteffekt S_H beschreibt die Interaktionseffekte. Die restlichen Eingänge sind für den Modellausgang \dot{m}_{NO_x} nahezu irrelevant ($S_i < 0,05$). Die Längenskalare des GPMs liefern zu den varianzbasierten Methoden konsistente Ergebnisse. Die Haupteinflussfaktoren AGR-Rate ψ_{AGR} und EZMI ϕ_{MI} sind in veränderter Reihenfolge signifikant. Im Fall des effektiven Mitteldruck-Modells p_{eff} bestimmt der Haupteinfluss EZMI ϕ_{MI}

den Modellausgang (Abb. 3.10b). Die nichtlinearen Zusammenhänge der restlichen Eingangsparameter sind für den Modellausgang weitgehend unwesentlich. Der Signaleingang p_{ALD} steht an zweiter Stelle der varianzbasierten Methode $S_i = 0,17$. Folglich genügen statistische Zusammenhänge der Datensätze für Aussagen über die Modellsensitivität nicht.

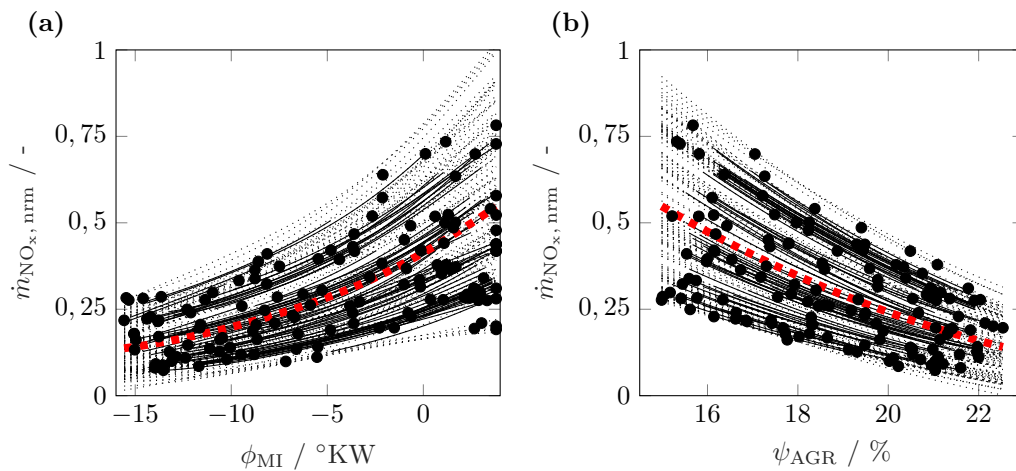


Abbildung 3.11 PDP (rot) und ICE-Plot (schwarz) für die Haupteinflussfaktoren – den EZMI ϕ_{MI} (a) und die AGR-Rate ψ_{AGR} (b) – im gesamten Faktori-Variationsbereich auf die Ausgangsgröße \dot{m}_{NO_x} . Extrapolierte Versuchsraumverletzungen der Faktorkombinationen sind gepunktet, gültige Modellauswertungen als durchgezogene Linie und die PD gestrichelt dargestellt

Die Gleichverteilung des Stichprobenumfangs im Hyperkubus der Schätzung wurde bereits erwähnt. Nachteilig berücksichtigt die Sensitivitätsanalyse in beschränkten Versuchsräumen Modellextrapolation in der Varianzzerlegung. Diese Verletzung der Modellgrenzen kann mit den beiden visuellen Erklärbarkeitsansätze PDP und ICE sowie der konvexen Hülle des Modells überprüft werden. In Abb. 3.11 sind die grafischen Einflüsse des Haupteffektes EZMI ϕ_{MI} im gesamten Variationsbereich auf das Modell aufgezeigt. Dabei bildet der PDP (rot) die durchschnittliche Modellvorhersage und die ICE-Linien (schwarz) die Modellauswertung bei der Variation der anderen Modelleingänge. Die schwarzen Punkte kennzeichnen Messdaten, die gepunkteten Linien markieren extrapolierte Variationen außerhalb der konvexen Hülle. Zusätzlich zur Sensitivität können mit Hilfe der grafischen PDP-Auswertung der monotone Zusammenhang und die Richtung der Einflussgröße bestimmt werden. Die ICE-Grafik kann mit sich überschneidenden Kurven Interaktionseffekte zwischen verschiedenen Kalibrierparametern darstellen.

Die Erklärbarkeitsansätze zeigen für das Beispiel der NO_x -Emission in Abb. 3.11a die Auswirkung einer Verlagerung des EZMIs ϕ_{MI} . Die NO_x -Emission sinkt für die zeitliche Verzögerung der Einspritzung. Das NO_x -Modell hat nahezu keine Interaktionseffekte, da fast alle Linien alleinstehen.

Eine weitere NO_x-Reduktion kann mit einer höheren AGR-Rate ψ_{AGR} erzielt werden (vgl. Abb. 3.11b). Die Sensitivitätsanalyse mittels PDP bzw. ICE kann für komplexe Anwendungsfälle physikalischer Vorgänge herangezogen werden und Effekte von Eingangs- auf Ausgangsgrößen sowie deren Effekte untereinander erklären.

3.3 Methoden der Totzeitschätzung

In der modellbasierten, transienten Kalibrierung von Motorsteuergeräten ist die TDE bei der Identifikation von dynamischen Modellen entscheidend. Totzeiten entstehen durch zeitbeanspruchende Prozesse [32]. In der Emissionsgesetzgebung ist die TDE für die Ausrichtung von aufgenommenen Gaskonzentrationen, Abgasmassenströmen, Fahrgeschwindigkeiten und sonstigen relevanten Signalen vorgeschrieben [2]. Der Fokus liegt auf der zeitinvarianten TDE.

Die Totzeit kann in *scheinbare* und *wahre* Zeitverzögerung unterschieden werden. Die *scheinbare* Verzögerung bezeichnet die Totzeit, die in der Modellidentifikation die bestmögliche Modellqualität ergibt [120, 11, 4, S. 26]. Die *scheinbare* Zeitverzögerung kann sich für komplexe Modellstrukturen einstellen, obwohl der Prozess gar keine Zeitverzögerung hat. Das System schätzt die Totzeit fehlerhaft, da eine zu hohe Modellordnung angenommen ist. Die *wahre* Zeitverzögerung bezeichnet die reine Zeitverzögerung des physikalischen Prozesses mit minimalen Annahmen über die Prozessdynamik [120]. Die TDE mit mehreren Modellen ergibt sich bei dem Modell mit der höchsten Modellqualität. Dies muss nicht zwingend der *wahren* Totzeit entsprechen [72]. Im Rahmen dieser Arbeit liegt der Fokus auf der Schätzung der *wahren* Zeitverzögerung. Einen breiten Überblick über Methoden der TDE gibt [11]. Die Schätzmethoden der Totzeit sind in die **Schätzung im Zeitbereich** in Kap. 3.3.1 und die **explizite Parameterschätzung** in Kap. 3.3.2 eingeteilt, angelehnt an Björklund] [11].

3.3.1 Schätzung im Zeitbereich

Die Ähnlichkeit zweier Signale $u(k)$ und $y(k)$ misst die Kreuzkorrelationsfunktion in Abhängigkeit vom zeitlichen Schiebeoperator κ . Die Funktion verschiebt und multipliziert die Signale $u(k)$ und $y(k + \kappa)$. Das resultierende Produkt wird integriert und über das Zeitintervall normalisiert. Der maximale Wert der Kreuzkorrelationsfunktion $|r_{yu}(\kappa)|$ gibt die höchste Ähnlichkeit beider Signale bei der Totzeit n_k an [63, S. 20]. Die Kreuzkorrelationsfunktion in Gl. 3.23 ist nutzbar bei linearen Systemen [9, S. 120]. Im Bereich großer Nichtlinearität kann die Kreuzkorrelationsfunktion scheitern, nichtlineare Effekte zu erkennen [9, S. 120].

$$r_{yu}(\kappa) = \mathbb{E}\{U(k) \cdot Y(k - \kappa)\} = \lim_{K \rightarrow \infty} \frac{1}{2K + 1} \sum_{k=-K}^K u(k)y(k - \kappa) \quad (3.23)$$

Eine Verschiebung des Signals $y(k)$ mit dem Schiebeoperator κ kann auf Methoden der Datenanalyse angewendet werden. Für jede Verschiebung wird ein Korrelationskoeffizient nach *Pearson* mit der Gl. 3.1 berechnet. Das Maximum sämtlicher Koeffizienten abhängig von den Verschiebungen κ gibt die Totzeit n_k wieder. Sowohl die Maximierung des Korrelationskoeffizienten als auch die Kreuzkorrelationsfunktion sind von der Europäischen Union für die Berechnung von Totzeiten vorgeschrieben [2]. Da beide Ansätze für lineare Zusammenhänge gelten, wird die Schätzung der Totzeit mittels der Transinformation vorgeschlagen [36, 77]. Die Totzeit n_k kann in nichtlinearen Datensätzen durch die Minimierung der Entropie oder umgekehrt durch die Maximierung der Transinformation bestimmt werden. Exemplarisch ist die Funktion der verallgemeinerten Transinformation in Abhängigkeit vom Schiebeoperator κ in Gl. 3.24 gegeben.

$$I_2(U; Y(\kappa)) = \sum_{i=1}^N \sum_{j=1}^M p_{UY}(x^{(i)}, y^{(j+\kappa)}) \text{ld} \frac{p_{UY}(x^{(i)}, y^{(j+\kappa)})}{p_U(x^{(i)})p_Y(y^{(j+\kappa)})} \quad (3.24)$$

3.3.2 Explizite Parameterschätzung

In Kap. 2.2 werden die nichtlinearen dynamischen Modelle NARX, NOE und NFIR vorgestellt. Deren lineare Äquivalente *Autoregressive with exogenous input* (ARX) und *Output error* (OE) sowie *Autoregressive moving average with exogenous input* (ARMAX) sind in Abb. 3.12 dargestellt [85, S. 558, 12, S. 291]. Die explizite Parameterschätzung bildet für einen ganzen Satz an Verzögerungen jeweils ein lineares dynamisches Modell mit einer zugehörigen Modellqualität. Das Modell mit der höchsten Genauigkeit resultiert bei der Verzögerung, welche die *wahre* oder die *scheinbare* Totzeit wiedergibt [72]. Bei der Wahl einer falschen Modellstruktur ist die Bestimmung der *wahren* Totzeit fehlerbehaftet, da das Modell mit dem niedrigsten Modellfehler die *scheinbare* Verzögerung ermittelt [11]. Die TDE mit nichtlinearen dynamischen Modellstrukturen ist aufgrund des im Vergleich deutlich erhöhten Rechenaufwands der numerischen Lösung nicht aufgeführt.

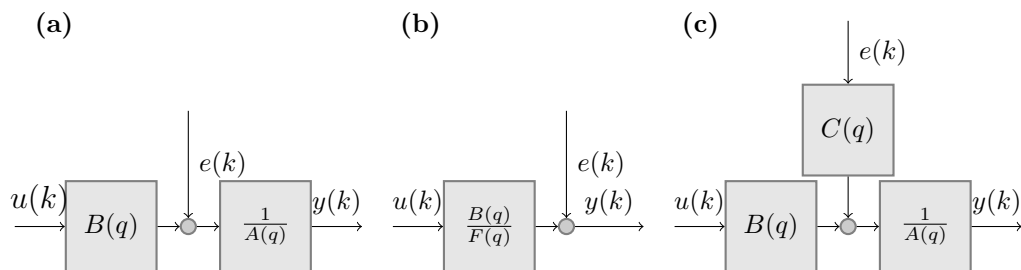


Abbildung 3.12 Lineare dynamische Modellstrukturen für die explizite Parameterschätzung der Totzeiten: ARX- (a), OE- (b) und das ARMAX-Modell (c)

ARX-Modell

Das ARX-Modell ist in Abb. 3.12a und in Gl. 3.25 dargestellt. Die explizite Parameterschätzung berechnet für einen Vektor an Verzögerungen \mathbf{n}_k jeweils ein separates ARX-Modell für jeden einzelnen Vektoreintrag von \mathbf{n}_k . Die Verlustfunktion der Systemidentifikation bildet sich aus der Differenz des gemessenen Ausgangssignals eines Systems und des mit einem Modell berechneten Signals [12, S. 245]. Als Fehlerfunktion wird der *mittlere Quadratische Fehler* (MSE) verwendet. Die Parameterschätzung des linearen Optimierungsproblems eines ARX-Modells erfolgt mit der LS-Methode. Der Vorteil der TDE mit zahlreichen ARX-Modellen liegt in dem geringen numerischen Aufwand im Vergleich zu komplexen Modellstrukturen [72, S. 8, 85]. Das ARX-Modell

$$A(q)y(k) = B(q)u(k - n_k) + e(k) \quad (3.25)$$

ist mit den charakteristischen Polynomen $A(q)$ und $B(q)$ sowie dem externen Stör-signal $e(k)$ definiert [72, S. 8, 85]. Die Ordnung entspricht der Anzahl der zurückliegenden Zeitschritte, die den Eingang und den Ausgang beeinflussen. Die Ordnungen für die Polynome der Ein- und Ausgangsdaten n_a und n_b werden entweder vor der Parameterschätzung festgelegt oder durch die Aufteilung der Messdatensätze in Identifikations- und Validierungsdaten ermittelt. Aus den Identifikationsdaten wird für sämtliche Kombinationen an Ordnungen jeweils ein Modell geschätzt. Die geschätzten Modelle werden mit den Validierungsdaten überprüft, wobei der geringste Modellfehler die Ordnungen n_a , n_b sowie die *scheinbare* Totzeit n_k wiedergibt. Ein überkomplexes Modell mit zu hohen Modellordnungen kann im Pol-Nullstellen-Diagramm überprüft werden. Kompensieren die Nullstellen der geschätzten Transferfunktion genau die Anzahl o_m der Pole, ist die Modellordnung n_m zu groß gewählt und die Modellordnung des Systems entspricht $n_m - o_m$ [85]. Speziell im Fall der *wahren* TDE mit ARX-Modellen werden für die Zähler- $B(q)$ und Nennerpolynome $A(q)$ Modellstrukturen zweiter bzw. vierter Ordnung empfohlen [28, 72].

Das Nennerpolynom $A(q)$ beschreibt sowohl die Systemdynamik als auch die Prozessstörung (siehe Abb. 3.12a). Aus diesem Grund sollte das ARX-Modell bei Messdaten mit großem Signal-Rausch-Abstand verwendet werden [85, S. 629]. Das ARX-Modell beruht auf der Annahme, dass der abzubildende reale Prozess nicht gestört ist. Im nichtgestörten Fall beinhalten die Parameter einen systematischen Modellfehler und sind nicht konsistent. Der systematischen Modellfehler (Bias) umfasst den Unterschied zwischen dem geschätzten Modell und dem wahren Prozessverhalten ohne Rauschen. Die Verwendung von *Instrumentenvariablen* (IV) in der Modellschätzung bietet eine Möglichkeit, um die Konsistenzprobleme von ARX-Modellen in in stark gestörten Prozessen zu verhindern und den systematischen Modellfehler zu verkleinern [85, S. 485].

OE-Modell

Eine weitere Möglichkeit, die Ungenauigkeiten in der ARX-Schätzung aufzufangen, ist die Nutzung von komplexen Modellstrukturen für die TDE. Praktische Verwendung finden das Ausgangsfehlermodell (OE-Modell) oder das ARMAX-Modell [11]. Die Systemidentifikation mit komplexen Modellen resultiert in höheren Systemgenauigkeiten und in Totzeiten mit höherer Genauigkeit [28]. Bei der ARMAX-Modellstruktur wird die ARX-Struktur um einen Term erweitert, der vergangene Werte der Störgrößen mitberücksichtigt (siehe Abb. 3.12c). Im Vergleich zu ARX- und ARMAX-Modellen ist beim OE-Modell die Systemdynamik unabhängig vom Störmodell (siehe Abb. 3.12b). Das Störsignal folgt im Prozess additiv. Da das Störmodell ohne Filterung durch das Polynom $A(q)$ aufgebaut ist, erfolgt die Parameterschätzung des OE-Modells mit nichtlinearen Methoden [85]. Durch die fehlende Störtransferfunktion sind OE-Modelle empfindlich in der Parameterschätzung gegenüber niedrigen Frequenzen [12]. Das OE-Modell ist beschrieben mit [85, S. 468]

$$y(k) = \frac{B(q)}{F(q)}u(k - n_k) + e(k). \quad (3.26)$$

Algorithmus eines vereinfachten ARMAX-Modells

Der Vorteil des OE-Modells gegenüber einer ARX-Struktur liegt in der höheren Genauigkeit. Sein Nachteil liegt in der Parameterschätzung des nichtlinearen Optimierungsproblem und dem damit verbundenen rechnerischen Aufwand. [11] kombiniert die Stärken der beiden Ansätze zu der gefilterten Methode MET1STRUC. Der Algorithmus ist in die folgenden vier Schritte unterteilt:

1. Parameterschätzung eines Zustandsraummodells, Gl. 3.27
2. Umwandlung der Systemstruktur in eine ARMAX-Modellstruktur Gl. 3.28
3. Filterung des Eingangs $u(k)$ und Ausgangs $y(k)$ mit dem Polynom $1/C(q)$
4. Schätzung einer Schar von ARX-Modellen nach Gl. 3.25 mit jeweils unterschiedlichen Totzeiten und anschließender Vergleich der Modellgüte

Im ersten Schritt wird ein Zustandsraummodell geschätzt. Das geschätzte Zustandsraummodell kann anschließend in eine ARMAX-Modellstruktur (Abb. 3.12a) umgewandelt werden, sofern die Ordnungen der Polynome $A(q)$, $B(q)$ und $C(q)$ für die Systembeschreibung hoch genug gewählt sind [120, S. 657, 11]. Das Zustandsraummodell ist mit der Systemmatrix \mathbf{A} , dem Ausgangsvektor \mathbf{c}^T , dem Durchgriff d , dem Eingangsvektor \mathbf{b} und dem Störsignal $e(k)$ definiert als [85, S. 505]

$$\begin{aligned} \dot{\mathbf{x}}(k) &= \mathbf{A} \mathbf{x}(k) + \mathbf{b}u(k) \\ y(k) &= \mathbf{c}^T \mathbf{x}(k) + du(k) + e(k). \end{aligned} \quad (3.27)$$

Die zeitliche Änderung des Zustandsvektors $\dot{\boldsymbol{x}}(k)$ ist dabei eine Funktion des momentanen Wertes des Zustandsvektors $\boldsymbol{x}(k)$ sowie der Eingangsgröße $u(k)$ [12, S. 239]. Eine direkte Parameterschätzung des ARMAX-Modells ist wegen des Polynoms $C(q)$ im Störmodell nichtlinear und numerisch aufwendig. Mit Schätzung eines Zustandsraummodells nach *Schritt 1*. verringert sich der Rechenaufwand. Das Modell wird mit der *State-space subspace system identification* (4SID) Methode geschätzt [132]. Eine ARMAX-Modellstruktur ähnelt dem Aufbau einer ARX-Struktur und ist in Gl. 3.28 gegeben [85, S. 494]. Das ARMAX-Modell besitzt die gleichen Polynome $A(q)$ und $B(q)$ für die Systemdynamik. Das Störmodell und wird mit dem Polynom $C(q)$ zur Beschreibung von Störungen ergänzt. Das Polynom $C(q)$ eignet sich bei dominierenden Störungen früh im Prozess [59, S. 168].

$$A(q)y(k) = B(q)u(k - n_k) + C(q)e(k). \quad (3.28)$$

3.3.3 Verifikation und Vergleich der Schätzmethoden

Die Methoden der TDE werden mittels zweier generischer Datensätze nachfolgend verglichen: einem linearen sowie einem nichtlinearen Fall. Speziell im nichtlinearen Fall liegen die Herausforderungen in einer korrekten Wahl der Schätzmethode, da einige Schätzmethoden für lineare Zusammenhänge gelten. Dafür wird ein vordefiniertes Zufallssignal $u(k)$ mit einer linearen Zustandsraummodellierung auf ein lineares Ausgangssignal $y_L(k)$ transformiert. Gleichermaßen wird das Zufallssignal $u(k)$ nichtlinear auf das Ausgangssignal $y_{NL}(k)$ transformiert. Die transformierten Signale $y_L(k)$ und $y_{NL}(k)$ werden anschließend auf eine vordefinierte Totzeit $n_k = 23$ verschoben. Diesen Wert gilt es von den Schätzmethoden der Totzeit exakt zu treffen. Anhand der Ergebnisse der Totzeit der jeweiligen Schätzmethoden erfolgt eine Empfehlung.

In Abb. 3.13 sind die drei generischen Signale $u(k)$, $y_L(k)$ und $y_{NL}(k)$ gegenübergestellt. Das Zufallssignal $u(k)$ ist im Intervall $[0, 20]$ gleichverteilt und mit einem linearen Trend behaftet. Die Zahlenwerte und Zustandsraummodelle der linearen (Gl. C.32) und nichtlinearen Transformation (Gl. C.34) sind im Anhang angegeben. Die Modellkomplexität der Parameterschätzung wird mit der maximalen Modellordnung $n_m = 8$ beschränkt, um den Rechenaufwand zu reduzieren.

Der statistische Zusammenhang der beiden Anwendungsbeispiele, linear und nichtlinear, lässt sich folgendermaßen in Zahlenwerten ausdrücken: Der Korrelationskoeffizient nach *Pearson* quantifiziert die Linearität der vorgestellten Ausgänge der Zustandsraumtransformationen zum Zufallssignal $u(k)$. Der Faktor bewertet den linearen Zusammenhang zwischen dem Eingangs- $u(k)$ und dem Ausgangssignal $y_L(k)$ mit $r_{P,L} = 0,4$. Der nichtlineare Fall ergibt einen Wert von $r_{P,NL} = 0,04$. Die Schätzwerte der verallgemeinerten Transinformation als Maß für den nichtlinearen statistischen Zusammenhang sind linear $\hat{I}_2(u; y_L) = 0,66$ sowie nichtlinear $\hat{I}_2(u; y_{NL}) = 0,17$.

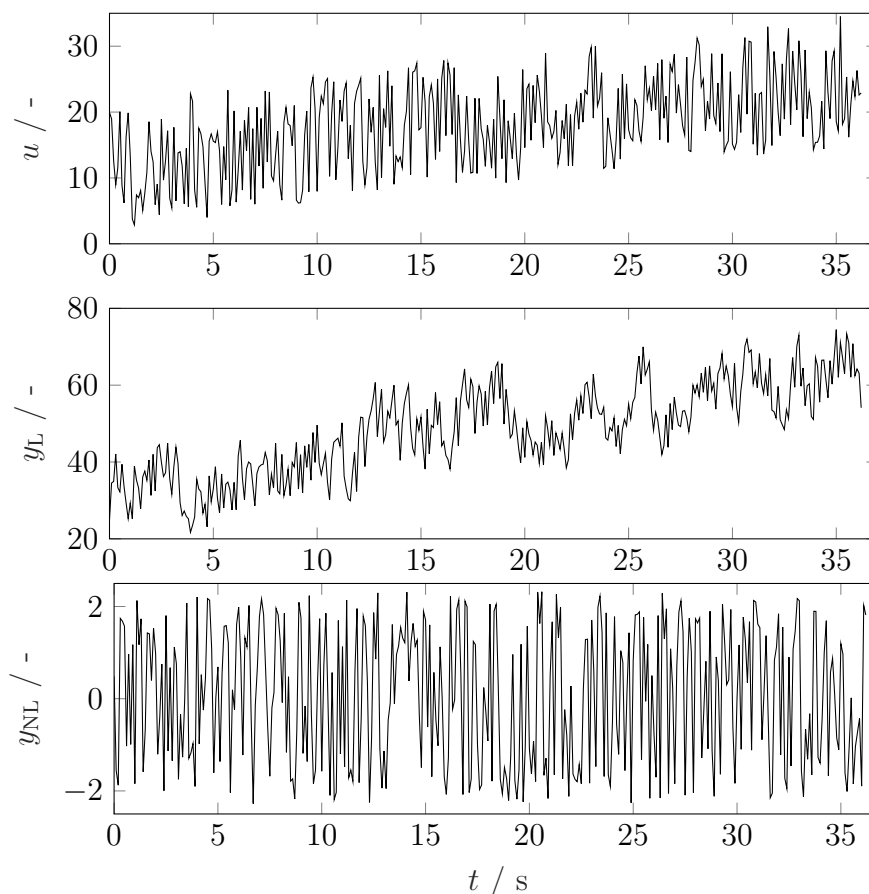


Abbildung 3.13 Vergleich der Signale zur TDE: Zufallssignal u , linear transformiertes Signal y_L und nichtlinear transformiertes Signal y_{NL}

Die Ergebnisse der Schätzmethoden im Zeitbereich sowie der Parameterschätzung der Totzeitglieder sind in Tab. 3.5 gezeigt. Die Verschiebung $n_k = 23$ der Signale $u(k)$ und $y_L(k)$ im linearen Fall wird bei sämtlichen Schätzmethoden exakt ermittelt. Die Kreuzkorrelationsfunktion abhängig vom Schiebeoperator κ ist in Abb. 3.14 dargestellt. Die Verschiebung des linearen Ausgangssignals mit dem Totzeitglied $y_L(k+n_k)$ maximiert den Korrelationskoeffizienten auf einen Wert von $r_{P, L} = 0,74$. Die Parameterschätzung mit den ARX-Modellen erfolgt mit geringer Modellordnung $n_m \leq 4$. Die Benutzung der IV reduziert die Ordnung auf den Wert $n_m = 1$. Die Schätzung mit der verallgemeinerten Transinformation ergibt die exakte Verzögerung bei der Totzeit $n_k = 23$ und dem maximalen Wert der Schätzung $\hat{I}_2(u; y_L) = 0,72$.

Die Herausforderungen bei der TDE liegen bei nichtlinearen Zusammenhängen der Datensätze. Im nichtlinearen Fall scheitern die Schätzungen der Korrelationsmethoden. Die Kreuzkorrelationsfunktion z. B. resultiert in einer akasalen negativen Totzeit. Das Ergebnis der Kreuzkorrelationsfunktion in Abhängigkeit vom Schiebeoperator κ ist in Abb. 3.14 gezeigt. Die hohe Anzahl an lokalen Minima und Maxima stellt bildlich die Nichtlinearität des Systems dar.

Tabelle 3.5 Zusammenfassung der Ordnungen und Ergebnisse der Schätzung der Totzeitglieder n_k der generischen Signale

Methode	linearer Fall			nichtlinearer Fall		
	n_k	n_a/ n_f	n_b	n_k	n_a/ n_f	n_b
Kreuzkorrelation r_{yu}	23			-47		
Korrelationskoeffizient r_P	23			148		
Transinformation \hat{I}_2	23			39		
ARX 2. Ordnung	23	2	2	34	2	2
ARX 4. Ordnung	23	4	4	34	4	4
ARX	23	3	1	47	1	1
ARX IV	23	1	1	47	1	1
MET1STRUC	23	3	1	37	7	1
OE-Modell mit $nf = 1$	23	1	2	34	1	1
OE-Modell mit $nf = 2$	23	2	2	34	2	4

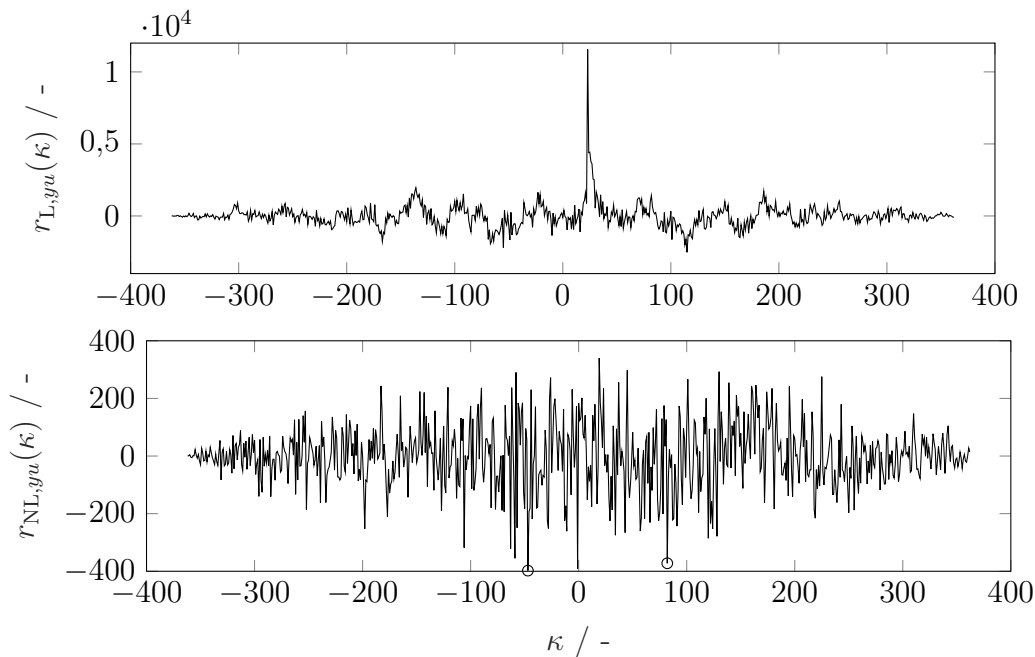


Abbildung 3.14 Kreuzkorrelationen der linearen und nichtlinearen Signale als Funktion des Schiebeoperators κ

Die Korrelationsfunktion schätzt die Verzögerung auf $n_k = -47$. Eine Totzeit von $n_k = 82$ folgt für eine Fensterung der Funktion auf den Bereich positiver Verschiebungen $\kappa = [0, 100]$. Der Wert der verallgemeinerten Transinformation ist höher als der des Korrelationskoeffizienten $r_{P,NL} = 0,13 < \hat{I}_2(u; y_{NL}) = 0,25$. Dies lässt auf einen statistischen Zusammenhang schließen.

Die Parameterschätzverfahren der OE- und ARX-Modelle 2. und 4. Ordnung überschätzen die exakte Totzeit. Die Rechenzeit der ARX-Parameterschätzung beträgt etwa 2% der Gesamtrechenzeit der numerischen Schätzung des OE-Modells in diesem Beispiel. Die Verwendung der IV bewirkt in beiden Anwendungsfällen keine Verbesserung der Schätzung.

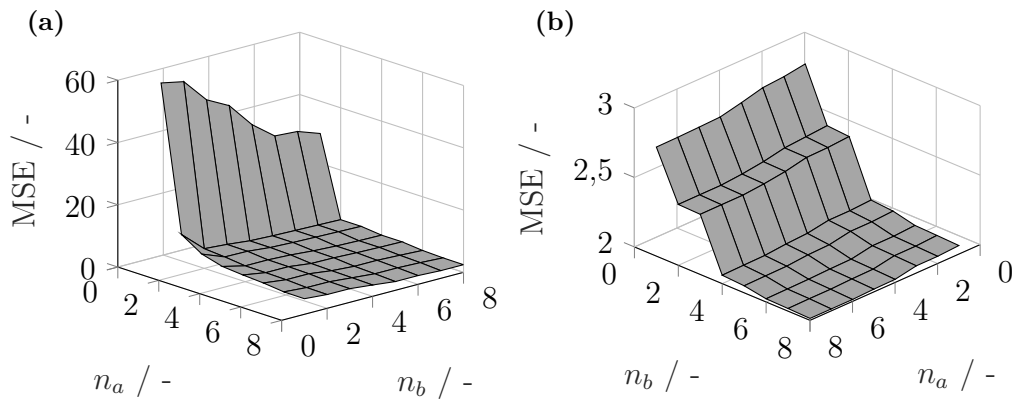


Abbildung 3.15 Vergleich des Modellfehlers MSE für ARX-Modellschätzungen bis zur Ordnung $n_m = 8$ für linearen (a) und nichtlinearen Anwendungsfall (b)

Bei der Parameterschätzung mit linearen dynamischen Modellen werden Modellfehler verglichen, die für eine steigende Modellordnung geschätzt werden. Für den linearen und nichtlinearen Fall sind die Modellfehler des ARX-Modellansatzes über den Polynomordnungen n_a und n_b bis zur Ordnung $n_m = 8$ in Abb. 3.15 aufgezeigt. Die *scheinbaren* Totzeit liegt im globalen Minimum aller Modellfehler. Die *wahre* Totzeit kann an der Stelle vermutet werden, bei der die Modellqualität ein Plateau erreicht. Eine weitere Erhöhung der Modellordnung bewirkt marginale Verbesserung im Modellfehler.

In Abb. 3.16 ist ein praktisches Beispiel der TDE von Messdaten exemplarisch am Abgasmassenstrom \dot{m}_{exh} und der Konzentration der Partikelanzahl c_{PN} gezeigt. Der Versuch einer TDE in Anwendungen mit Signalen der Partikelanzahl ist aufgrund des geringen linearen Zusammenhangs zu Signalen der Motorsteuerung besonders herausfordernd. Eine korrekte Schätzung ist für die Zertifizierung zwingend notwendig, um Konzentrationen zu Massenströmen umzuwandeln, und gesetzlich vorgeschrieben [2, 32]. Der Korrelationskoeffizient zwischen den beiden Signalen aus der Abb. 3.16 ergibt $r_{\text{P, PN}} = 0,02$. Die TDE mit der Kreuzkorrelationsfunktion scheitert in dieser Anwendung. Eine fehlerhafte Schätzung hat einen Einfluss auf das Bestehen einer Zertifizierung [32, 98].

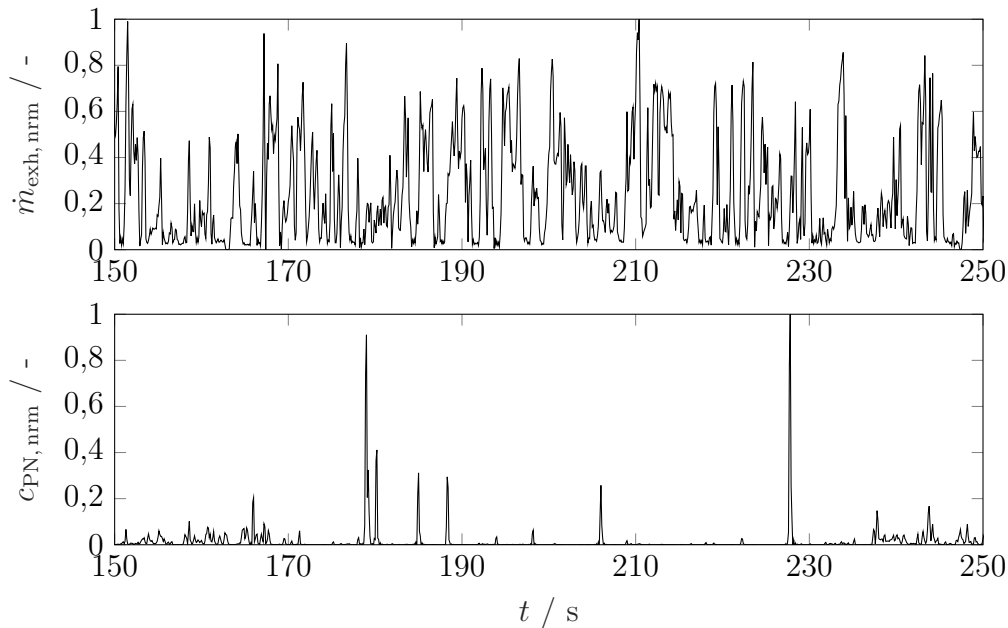


Abbildung 3.16 Ein Ausschnitt der normierten Signale des Abgasmassenstromes \dot{m}_{exh} und der Konzentration der Partikelanzahl c_{PN} für eine Berechnung des Abgaspartikel-Massenstroms in der Zertifizierung

3.4 Methodischer Ansatz zur RDE-Kalibrierung

Mit der Abgasgesetzgebung und der Ergänzung des Prüfverfahrens um RDE liegt das Ziel der Kalibrierung von Steuergeräteparametern in der sicheren Beherrschung der Emissionen des Antriebssystems unter realen Fahrbedingungen unabhängig von möglichen Fahreinflüssen. Dies stellt die Entwicklung vor eine große Herausforderung [3, S. 109]. Sind die Zielgrößen der Kalibrierung von Verbrennungsmotoren in (quasi-)stationären Betriebspunkten optimiert, führen hohe Lastgradienten zu einem dynamischen Betrieb und dadurch zu überdurchschnittlich hohen Schadstoffemissionen [131, S. 1275]. Eine Maßnahme zur Schadstoffoptimierung ist die Phlegmatisierung [131, S. 1275, 21]. Die dynamischen Anteile des Verbrennungsmotors werden im transienten Fahrbetrieb von einer elektrischen Maschine übernommen, welche die Leistungsdifferenz liefert. Die Höhe, Dauer und Dynamik der Ausgleichsgröße wirken sich auf die Haltbarkeit des elektrischen Speichers aus, wodurch mit Verschleiß zu rechnen ist [15, S. 321].

Der Schwerpunkt der Entwicklung von Ottomotoren liegt in der Verbrauchsreduktion [69, S. 3]. Beim Dieselmotor liegt der Fokus der Entwicklung auf der Einhaltung der Emissionsvorschriften [69, S. 3]. Ein Erreichen der Emissionsgrenzwerte ist alleine innermotorisch ohne modernste Abgasnachbehandlung nicht mehr zu realisieren [69, S. 72, 3, S. 112, 84, S. 1].

Eine applikative Optimierung der Schadstoffemissionen in der Dieselmotorentwicklung kann generell in drei Punkte gegliedert werden [3, S. 111]: die Optimierungen im stationären Betrieb, im transienten Betrieb oder im Betrieb unter bestimmten Fahrbedingungen: dies sind beispielsweise Hitze, Kälte oder Höhe. Ein Applikationsziel in der Prozessoptimierung des Dieselmotors sind u. a. die sich gegenläufig verhaltenden NO_x - und Partikelemissionen [69, S. 1, 84, S. 22]. Eine Zusammenfassung der Optimierungsziele der Motorenkalibrierung ist in [69, 84, 131, 97] zu finden. Diese umfassen die Senkung des Kraftstoffverbrauchs und der Emissionen unterhalb der legislativen Grenzwerte, minimale Geräuschemissionen, die Einhaltung bestimmter Bauteilbelastungsgrenzen und ein optimales Fahrverhalten. Der theoretische und experimentelle Fokus dieser Arbeit liegt auf dem Applikationsziel NO_x . Ein Bauteilschutz ist dahingehend gewährleistet, da sämtliche Versuchsumfänge eine entsprechende Sicherheit im Variationsbereich der Parameter berücksichtigen.

3.4.1 Werkzeugkette der modellbasierten Kalibrierung

Einen entscheidenden Beitrag zur Einhaltung der legislativen Emissionsgrenzen kann die modellbasierte Grundmotorenkalibrierung leisten. Das Potenzial folgt aus einem Vergleich von WLTC-Abgasrollentests verschiedener Fahrzeugklassen. In Abb. 3.17 sind kumulierte NO_x -Emissionsmassen über vier unterschiedliche Fahrzeugmassen dargestellt. Die Emissionskonzentrationen sind hinter der Abgasturbine (Rohemission) sowie nach den Katalysatoren (engl. *End of pipe* EOP) gemessen. Die Konzentrationen werden mit dem Abgasmassenstrom multipliziert, kumuliert und auf die Strecke bezogen. In Abb. 3.17a ist ein eindeutig linearer Zusammenhang der Fahrzeugmassen m auf die NO_x -Rohemissionsmassen pro Strecke m_{NO_x}/s festzustellen. Der lineare Korrelationskoeffizienten ergibt fast eins ($r_P = 0,97$). Die EOP-Emissionen in Abb. 3.17b hingegen lassen einen schwachen linearen statistischen Einfluss auf die Fahrzeugmasse erkennen ($r_P = 0,36$). Der Unterschied zum Korrelationskoeffizienten der Rohemissionen liegt in der Wirkung der Abgasnachbehandlung. Die Dimensionierung der Abgasanlage kommt einer Filterung der dynamischen Vorgänge gleich. Somit bedingt ein großes Potenzial für die Verminderung der NO_x -Emissionen die Fokussierung auf die Rohemissionen.

Die Kernpunkte der in der Vergangenheit bewährten Werkzeugkette der modellbasierten Kalibrierung für die Zyklusvorhersage bilden:

1. Reduktion des Testzyklus auf eine bestimmte Anzahl von zu untersuchenden stationären Betriebspunkten, die durch das Wertepaar Drehzahl und Lastgröße definiert sind
2. Variation der zu kalibrierenden bzw. zu optimierenden Parameter an den unter 1. ermittelten Betriebspunkten experimentell. Messung der Zielgrößen, wie beispielsweise des Verbrauchs und der Rohemissionen

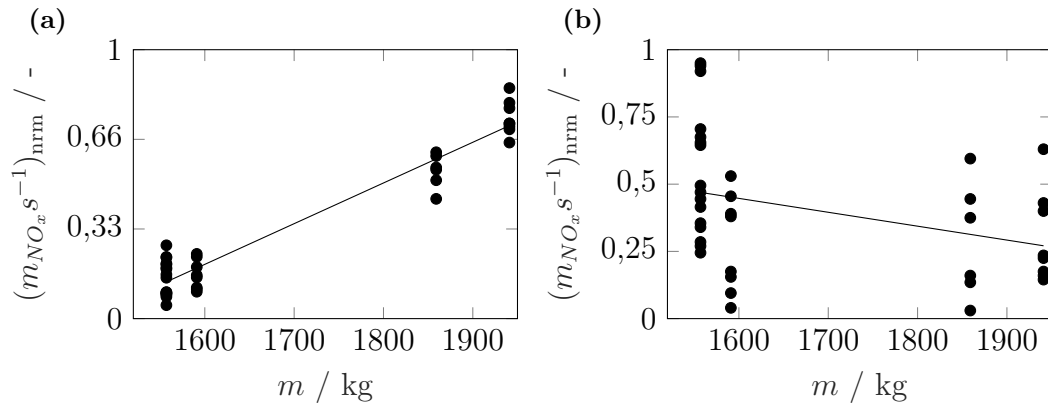


Abbildung 3.17 Einfluss der Fahrzeugmasse pro Strecke s im WLTC auf die NO_x -Roh- (a) und EOP-Emissionen (b) normiert

3. Modellierung der Zielgrößen in Abhängigkeit von den zu kalibrierenden Parametern mit einem linearen oder nichtlinearen Regressionsmodell
4. Minimierung des Verbrauchs unter Berücksichtigung der legislativen Emissionsgrenzen als Nebenbedingungen mit Hilfe eines numerischen Optimierungsansatzes für eine Zyklusvorhersage

Ziel der Prozesskette ist es, mittels modellbasierter Optimierung der Zielgrößen auf Basis eines Zertifizierungszyklus eine Zyklusvorhersage zu treffen. Der Zyklus wird im ersten Schritt auf ein statisches Optimierungsproblem reduziert. Zu Beginn werden durch einen *k-means*-Algorithmus die Betriebspunkte des transienten Prüflaufes für die Versuchsplanung gewählt, definiert durch die Drehzahl-Last-Kennfeldstruktur der ECU [68]. Mittels einer Versuchsplanung werden die zu kalibrierenden Parameter im Versuchsraum variiert, welche am Motorprüfstand vermessen werden. Nach der Modellidentifikation werden per Optimierungsverfahren die Parameter festgelegt. Die Validierung der neu parametrisierten Steuergerätefunktionen erfolgt mit dem eingangs vermessenen Zertifizierungszyklus auf die Zielgrößen. Somit kann für eine bestimmte Parametrierung eine Zyklusvorhersage modellbasiert bestimmt werden.

Die Reduktion des Zyklus auf Stationärpunkte durch ein Cluster-Verfahren ist auf die quasistationäre Charakteristik des NEDC zurückzuführen. Die Clusteranalyse teilt die Datenmenge mit einem *k-means*-Algorithmus in Segmente ein, deren Mittelpunkt einen sog. Cluster-Schwerpunkt darstellt. Die Schwerpunkte der Clusterelemente repräsentieren das gesamte Drehzahl-Last-Kennfeld, welches eine Bahnkurve des Zertifizierungszyklus abbildet. Die Anzahl der Cluster-Schwerpunkte ist entscheidend für die globale Gültigkeit von statischen Modellen im modellbasierten Kalibrierprozess des Betriebsbereichs des Zyklus [68].

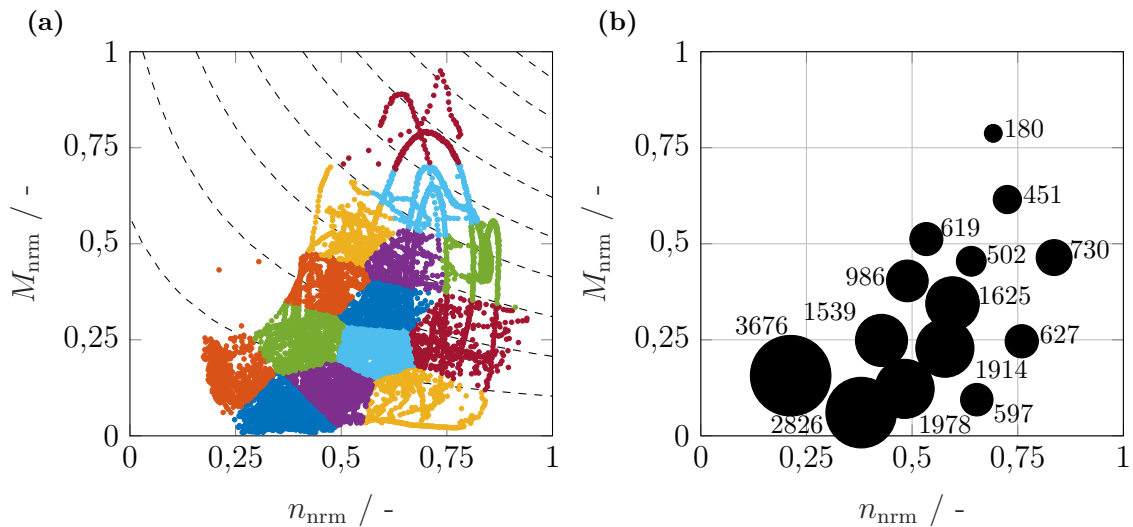


Abbildung 3.18 Ergebnisse einer Clusterung des WLTC im Drehzahl-Last-Kennfeld mit 14 Centroiden und Leistungshyperbeln (a) und der Verweildauer des Zyklus in den jeweiligen Centroiden (b)

Eine Umstellung der Prozedur auf den aktuellen Testzyklus WLTC lässt sich trotz der transienten Charakteristik verglichen mit dem NEDC einfach realisieren. Dafür wird die Anzahl der Cluster erhöht. Ein exemplarisches Ergebnis des *k-means*-Algorithmus eines WLTCs mit 14 Clustern ist in Abb. 3.18a gezeigt. Die Clusteranalyse behandelt einen um eine Randbedingung erweiterten *k-means*-Algorithmus, bei dem pro Segment die mittlere Leistung der segmentierten Einzelpunkte und des Schwerpunktes übereinstimmen [70, S. 6]. Die Verweildauer der Cluster-Schwerpunkte aus Abb. 3.18b kann mittels Multiplikation der jeweiligen Massenströme für eine Zyklusvorhersage verwendet werden.

Durch die RDE-Gesetzeserweiterung erhöht sich die Abdeckung des Betriebskennfeldes erheblich. In Abb. 3.19 sind Datenpunkte im Drehzahl-Last-Kennfeld des NEDC, einer Bahnkurve des WLTCs und einer Bahnkurve des RDE, jeweils mit deren Betriebsbereichen, gezeigt. Die Drehzahl bezieht sich im Folgenden auf die Motordrehzahl. Die Abdeckung des Kennfeldbereiches ist durch die Ablösung des NEDC durch den WLTC größer, verglichen mit dem abzudeckenden Bereich eines RDE-Ersatzzyklus allerdings gering. Die RDE-Bahnkurve weist eine höhere Abdeckung des Kennfeldes auf. Der erste Prozessschritt der Werkzeugkette für die modellbasierte Kalibrierung, die Clusteranalyse auf stationäre Betriebspunkte, basiert auf der Annahme einer quasistationären Charakteristik. Mit der gesetzlichen Erweiterung und der Einführung von RDE-Ersatzzyklen, die in der Zertifizierung neben dem WLTC betrachtet werden müssen, kann nach Abb. 3.19 und Abb. 3.20 nicht mehr von einem quasistationären Betriebsverhalten des Motors ausgegangen werden.

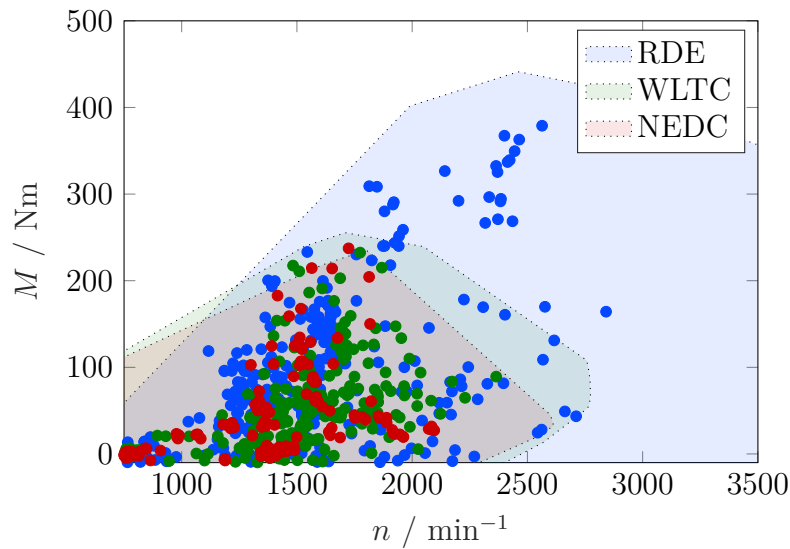


Abbildung 3.19 Abdeckung des Drehzahl-Last-Kennfeldes durch den synthetischen NEDC, den WLTC und einen möglichen RDE-Ersatzzyklus

Die Herausforderungen von RDE beschränken sich keineswegs auf den erhöhten Aufwand in der Grundlagenkalibrierung durch die erweiterte Abdeckung des Betriebskennfeldes. Vielmehr liegt die Komplexität in dem Wegfall der genormten, exakt reproduzierbaren Randbedingungen. Die Rahmenbedingungen sind bei WLTC und NEDC synthetisch klar vorgegeben. Die Logik der RDE-Einführung folgt der Einhaltung gesetzlicher Emissionslimits auf realen Straßenprofilen mittels mobiler Messtechnik und ist abhängig von Quereinflüssen und dem realen Fahrverhalten. Diese Erweiterung der Testbedingungen bringt erhöhte Anforderungen an die Systementwicklung und Fahrzeugabsicherung mit sich. Hersteller fällen nach der bisherigen Euro-Gesetzgebung Entscheidungen für Technologien auf Basis der reproduzierbaren Prüfzyklen NEDC und WLTC, welche eindeutig definiert sind und wovon im Wesentlichen die Thermodynamik bzw. Abgasnachbehandlung betroffen ist [78, S. 38]. Mit der Einführung von RDE erübrigen sich die normierten Kriterien für die Technologieentscheidungen. Eine geeignete Kalibrieremethode sollte möglichst unter den Randbedingungen des individuellen Fahrprofils und äußerer Rahmenbedingungen auf beliebige Ersatztestzyklen einsetzbar sein.

3.4.2 Aktueller Stand der Forschung zur Optimierung des transienten Betriebs

Die aktuelle Forschung beschäftigt sich mit Fragestellungen rund um die Optimierung des dynamischen Motorverhaltens und mit den kommenden Herausforderungen durch die Einführung von RDE [78, 84, 46, 130, 105]. Messtechnische Schwächen der Sensorik im Ansprechverhalten bei dynamischen Anwendungen behandeln die Publikationen [83, 56, 45]. Die jeweiligen Ansätze verfolgen eine Lösung mittels dynamischer Emissionsmodelle mit Totzeitkorrektur für die Abgaskomponenten NO_x und Ruß. [127] bewertet das Ansprechverhalten im transienten Fahrbetrieb von NO_x -Messtechniken und misst die charakteristischen Zeitgrößen. [5] zeigt Vorteile einer laserinduzierten Partikelmesstechnik für einen transienten Betrieb des Dieselmotors.

Einen Ansatz zur dynamischen Motorvermessung mit stochastischen Anregungssignalen für die stationäre Kennfeldoptimierung stellt Hafner [46] vor. Er entwickelt eine dynamische Vorsteuerung zweier Parameter, des Verhältnisses der Abgasrückführung und des Drucks der Turboaufladung, zur Verbesserung des Luft-Kraftstoff-Verhältnisses verbunden mit einer Senkung der dieselmotorischen Rußstöße. Dafür stellt er eine Simulationsumgebung auf und validiert diese in einer Prüfstandumgebung. Zydek [130] entwickelt zur Verhinderung transienter Emissionsbildung eine Reglerstruktur für den dynamischen Motorbetrieb und regelt einen optimalen Gaszustand unter dynamischen Bedingungen ein. Dazu werden die Stellgrößen Haupteinspritzwinkel, Raildruck und Drallklappenposition in Abhängigkeit vom Drehzahl-Last-Betriebspunkt und vom Gaszustand im Zylinder optimal gesteuert.

Maschmeyer [78] beschäftigt sich mit der Bewertung dynamischer Emissionskonformität und der Einflussbewertung dynamischer Motorbetriebsphasen für RDE. Es werden dynamische Betriebspunkttrampen am Prüfstand vermessen und auf einen erhöhten Emissionsausstoß hingewiesen. Dabei werden Drehmoment- und Drehzahlrampen berücksichtigt. Diese werden durch eine Differenz der Emissionsmassenströme zu deren quasistationären Verläufen optimiert [78, S. 109]. In der Versuchplanung wird zur Anregung der Dynamik ein maximaler Lastgradient als Sicherheitsbegrenzung über den Pedalwert η vorgegeben. Der maximale Lastgradient liegt bei $\dot{M} = d\eta/dt = \pm 50\% \text{ s}^{-1}$ [78, S. 108]. Auf die Problematik der Signalzuordnung bei Verzugszeiten ist hingewiesen. Die Lösung der TDE erfolgt durch die Zuordnung der Maxima der beiden Signalgradienten, des Drehmoments und der NO_x -Konzentration, bei einer Unterbrechung der Einspritzung [78, S. 115].

Naumann [84, S. 41] untersucht dynamische Motorbetriebsphasen und beschreibt den Einfluss von dynamischen Beschreibungsgrößen in der transienten Kalibrierung. Zur Verbesserung wird eine Erweiterung des Definitionsraums der Führungsgrößen-Kennfelder um eine dynamische Beschreibungsgröße empfohlen. Das Potenzial der Zylinderwandtemperatur und des Lastgradienten wird experimentell aufgezeigt sowie die deutliche Korrelation der beiden Größen beschrieben. Auf hohe Lastgradienten folgen starke Temperaturdifferenzen [84, S. 43].

Diese wissenschaftsbasierte Grundlagenuntersuchung betrachtet die gegenläufige Wechselbeziehung zwischen Kraftstoffverbrauch und NO_x -Emission im Lastpunkt $n = 1874 \text{ min}^{-1}$ und $\dot{M} = \pm 270 \text{ Nm s}^{-1}$ mit sechs Kennfeldvariationen des EZMI und zwei verschiedenen VTG-Steuerstrategien. Bei negativen Lastgradienten kann kein Rußstoß bei schnellen Lastsprüngen festgestellt werden [84, S. 49]. Die Arbeit folgt jedoch anschließend einer zyklusorientierten, quasistationären Kennfeldoptimierungsstrategie anstelle einer um die Dynamik erweiterten Kalibrieremethode, trotz des aufgezeigten Potenzials.

Einen Ansatz zur Optimierung transienter Fahrspuren und die Einführung von Kriterien für die Optimierungsgrößen behandelt Scheidel u. a. [105]. In dem Artikel werden Integrale und Maxima als skalare Ersatzgrößen der dynamischen Emissionsmassenströme gebildet, auf die eine stationäre Modellierung und Optimierung erfolgt. Die Versuchsplanung untersucht eine skalare Kennlinienverschiebung zur Ursprungskalibrierung für vier Einflussgrößen und eine sog. Drei-Punkte-Kennlinienerstellung für weitere zwei Parameter. Die skalare Kennlinienverschiebung multipliziert die parametrisierte Ursprungskalibrierung mit einem skalaren Wert, der variiert wird. Bei der Drei-Punkte-Kennlinienerstellung werden aus drei variierten Punkten jeweils Polynome gebildet. Die Variation der drei Punkte führt somit zu linearen oder quadratischen Kennlinien. Der Versuch beschränkt sich einzig auf den Lastsprung auf $\eta = 100\%$ in der Zeitspanne $t \approx 0$. Mit der optimierten Parametrierung für diesen maximalen Lastsprung wird über eine transiente Ansteuerung eine Rußeinsparung von 40% im WLTC realisiert [105, S. 227].

3.4.3 Erweiterung des Betriebspunkts um eine dynamische Führungsgröße

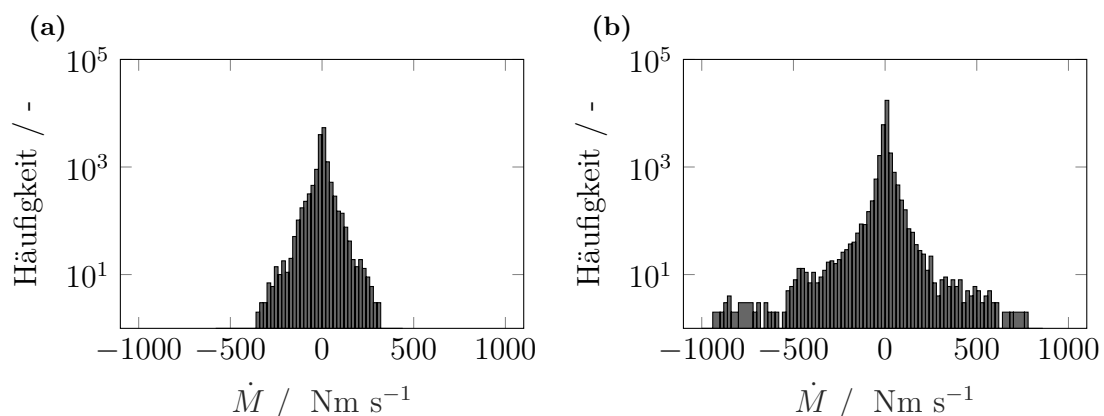


Abbildung 3.20 Häufigkeitsverteilung für den Lastgradienten $\dot{M} = dM/dt$ eines WLTC (a) und eines RDE-Ersatzzyklus (b)

In Abb. 3.19 ist die zweidimensionale Kennfeldabdeckung der beiden Führungsgrößen Drehzahl und Last dargestellt. Eine höhere transiente Ausprägung kann für den RDE-Ersatztest im Vergleich mit dem WLTC festgestellt werden, wenn die Dynamik des Drehmoments betrachtet wird. Der Unterschied ist mit einer Häufigkeitsverteilung des Lastgradienten $\dot{M} = dM/dt$ für beide Zyklen in Abb. 3.20 zu sehen. Die Spitzen der Häufigkeitsverteilung treten in beiden Fällen jeweils für die stationär betrachteten Punkte $\dot{M} \approx 0$ auf. Die Verteilung des RDE-Ersatzzyklus weist einen breiteren Wertebereich des Gradienten \dot{M} sowie eine höhere kumulierte Häufigkeit auf. Die höheren $\dot{M} > 200 \text{ Nm s}^{-1}$ und niedrigeren Wertespektren $\dot{M} < -200 \text{ Nm s}^{-1}$ sowie die breiteren Flanken der Verteilung des RDE-Ersatzzyklus belegen eine höhere dynamische Charakteristik verglichen mit der WLTC-Häufigkeitsverteilung. Folglich kann die Reduktion der transienten Fahrspur des RDE-Ersatzzyklus durch eine Clusteranalyse auf (quasi-)stationäre Betriebspunkte bedingt zu einer Optimierung des transienten Fahrverhaltens führen. Ein optimaler Motorbetrieb ist durch die Abweichung von dem stationär optimierten Zustand nicht mehr gegeben (vergl. [84, S. 41]).

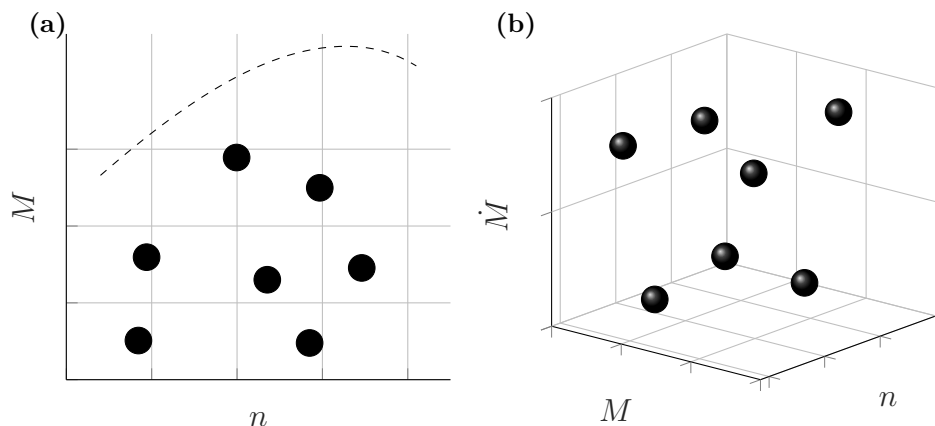


Abbildung 3.21 Exemplarischer Vorschlag einer neuen Definition des Betriebspunkts von einem zweidimensionalen Drehzahl-Last-Kennfeld (a) zu einem Wertetriple mit der Dynamik der ersten Ordnung, dem Lastgradienten, als dynamische Führungsgröße (b)

Die transiente Natur etwaiger RDE-Ersatzzyklen soll methodisch durch eine Erweiterung des Betriebspunkts um die Dynamik aufgefangen werden. Folglich definiert sich ein abgeleiteter Lösungsvorschlag für Prozessschritt eins der Werkzeugkette. Anstatt der zweidimensionalen Clusteranalyse des Führungsgrößen-Wertepaares Drehzahl/Last wird eine dreidimensionale, um eine dynamische Größe erweiterte Kennraumstruktur segmentiert. Dies ist exemplarisch in Abb. 3.21 dargestellt. Aufgrund der quasistationären Eigenschaft des NEDC ist die bisherige Betrachtung in der Grundmotorenkalibrierung vom Drehzahl-Last-Betriebskennfeld ausgegangen (vgl. Abb. 3.21a).

Eine zukünftige, dreidimensionale Clusteranalyse der RDE-Ersatztestzyklen berücksichtigt die Höhe der Dynamik des betrachteten transienten Testlaufes in einer dynamischen Größe. Ein mögliches dreidimensional erweitertes Führungsgrößen-Wertetripel ist an dem Beispiel des Lastgradienten \dot{M} in Abb. 3.21b skizziert. Aus der Formulierung des Applikationsziels folgt die Erweiterung der Steuergerätfunktionsstruktur zum Wertetripel. Der herkömmliche Kalibrierprozess in Abhängigkeit von den Kennfeld-Führungsgrößen Drehzahl n und Last M sowie deren aufgespannte Kennfeldwerte im Eingangs- bzw. Verstellparametervektor \mathbf{u} wird um den Lastgradienten \dot{M} erweitert. Der Lastgradient kann im Motorsteuergerät aus der momentanen Lastgröße, der Einspritzmenge oder dem Drehmoment, gebildet werden [84, S. 44].

$$\mathbf{y} = \mathbf{f}(n, M, \mathbf{u}) \quad \Rightarrow \quad \mathbf{y} = \mathbf{f}(n, M, \dot{M}, \mathbf{u}) \quad \text{mit} \quad \dot{M} = \frac{dM}{dt}. \quad (3.29)$$

3.4.4 Potenzialabschätzung der vorgeschlagenen Erweiterung

In Abb. 3.22 ist ein rohseitiges NO_x -Kennfeld eines Dieselmotors dargestellt, welches aus der quasistationären Datenwolke einer Drehzahl-Last-Lastgradient-Bahnkurve und dem Lastgradienten-Schnitt $\dot{M} = [-4, 4]$ eines RDE-Ersatzzyklus interpoliert ist. Die gesamte Datenwolke ist in Abb. 3.23 gezeigt. Aus der Kontur des interpolierten NO_x -Kennfeldes Abb. 3.22 geht ein stärkerer Effekt der Last M auf die NO_x -Entstehung aus als von der Drehzahl n . Die Korrelationskoeffizienten bekräftigen diese Beobachtung für die linearen statistischen Zusammenhänge von M und n auf den NO_x -Massenstrom \dot{m}_{NO_x} . Der Zusammenhang der Last $r_P = 0,84$ ist höher als der für die Drehzahl $r_P = 0,57$. Aufgrund des höheren Einflusses der Last auf die NO_x -Entstehung fällt die Wahl des zu untersuchenden Wertetripels auf den Lastanstelle des Drehzahlgradienten.

Die gesamte dreidimensionale NO_x -Bahnkurve des RDE-Ersatzzyklus ist in Abb. 3.23 gezeigt. In Punkten hoher Dynamik \dot{M} wird im Vergleich zum NO_x -Kennfeld aus den quasistationär geschätzten Datenpunkten eine höhere Emission vermutet, da mit der Dynamik zunehmend die stationär optimierten Kennfeldwerte nicht einer realen transienten optimalen Kalibrierung entsprechen. Gründe hierfür können ein verzögerter Ladedruck sein sowie abweichende thermische Zusammenhänge von Zustandsgrößen, die in der quasistationären Betrachtung nicht auftreten.

Der Einfluss der Dynamik auf die NO_x -Spur eines RDE-Ersatzzyklus kann mit einer Interpolation der Projektion der dreidimensionalen Bahnkurve geschätzt werden. Außerdem kann der Einfluss bspw. des Fahrprogramms analysiert werden. Dafür werden die NO_x -Emissionen der Bahnkurve im dreidimensionalem Betriebspunkt mit einer in das quasistationäre NO_x -Kennfeld aus Abb. 3.22 interpolierten Projektion verglichen. Die Interpolation der transienten Bahnkurve auf eine stationäre Projektion ist in Abb. 3.24 skizziert.

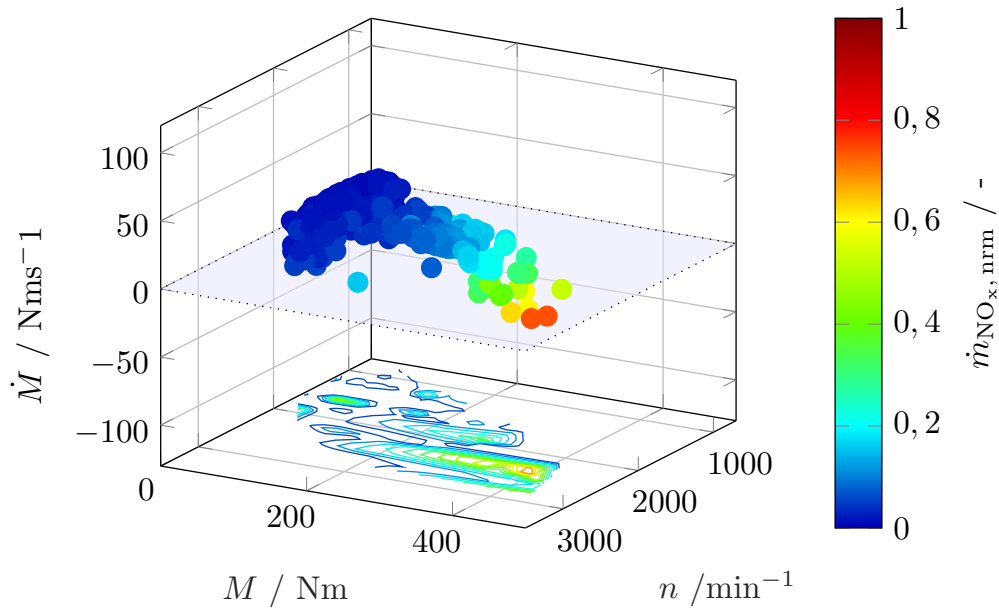


Abbildung 3.22 Interpolation eines Kennfeldes der normierten Rohemissionen von NO_x aus dem quasistationären Anteil $\dot{M} = [-4, 4]$ einer n - M -Bahnkurve eines RDE-Ersatzzyklus

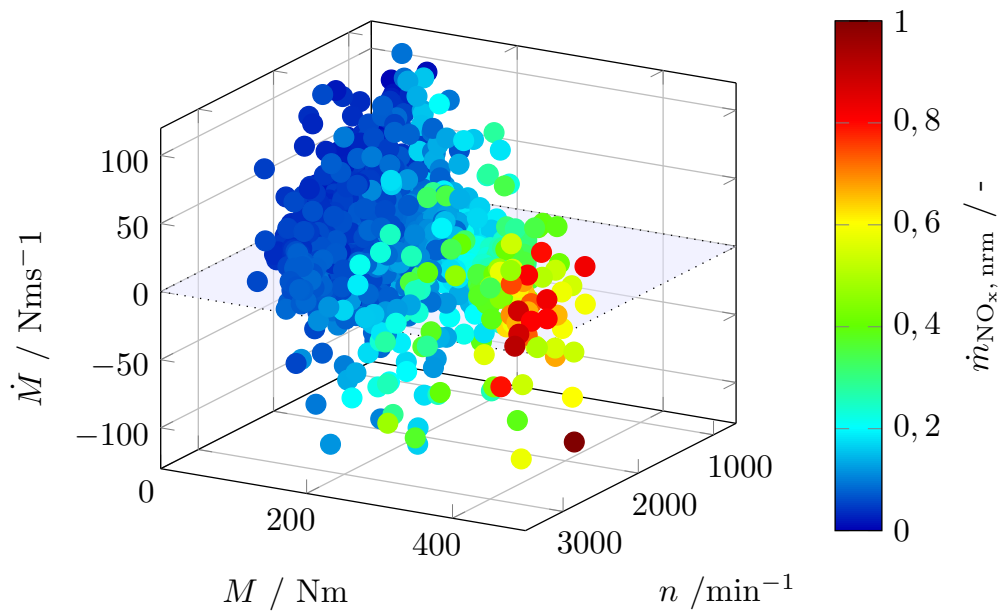


Abbildung 3.23 Normierte Rohemissionen von NO_x entlang der n - M - \dot{M} -Bahnkurve eines RDE-Zyklus und eine interpolierte Ebene mit dem quasistationären Anteil $\dot{M} = [-4, 4]$

Aus den kumulierten NO_x -Emissionen der interpolierten Projektion im Lastgradientenbereich $\dot{M} = [-4, 4]$ und den kumulierten realen NO_x -Emissionen der drei

dimensionalen Bahnkurve kann der Einfluss der Dynamik geschätzt werden. Die Differenz der kumulierten NO_x -Emissionen für die interpolierte quasistationäre Projektion der transienten Spur zu den realen Werten beschreibt das Einsparpotenzial unter der Annahme, dass die dreidimensionale Erweiterung des Definitionsraums der drei Führungsgrößen ($f(n, M, \dot{M})$) die dynamisch entstehenden Abweichungen zum stationären NO_x -Kennfeld auffangen kann.

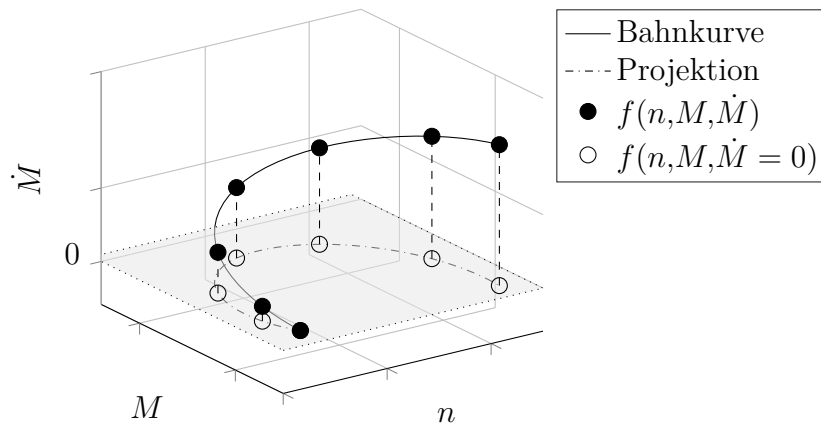


Abbildung 3.24 Prinzipskizze der Interpolation einer transienten $f(n, M, \dot{M})$ -Bahnkurve auf eine stationäre Projektion im Kennfeld $f(n, M)$

Tabelle 3.6 Ergebnisse der Potenzialabschätzung für die Erweiterung eines Betriebspunkts zu einem Wertetripel

Zyklus	Zusatz	Potenzial
WLTC		20 %
RDE 1	Fahrprogramm <i>Komfort</i>	28 %
RDE 1	Fahrprogramm <i>Sport</i>	31 %
RDE 2		37 %

In Tab. 3.6 sind die Abweichungen für verschiedene Arten von Fahrzyklen auf die NO_x -Zielgröße zusammengefasst. Kumulierte NO_x -Emissionen der zweidimensional interpolierten Projektion der dreidimensionalen Drehzahl-Last-Lastgradient-Bahnkurve sind geringer als die realen Emissionen der Bahnkurve. Es lassen sich daher Optimierungspotenziale bei einer Erweiterung des Definitionsraums der Führungsgrößen-Kennfelder um die dynamische Größe Lastgradient eines Zyklus für das Beispiel NO_x erkennen. Die prozentuale Höhe des Einsparpotenzials hängt von der Dynamik des Zyklus ab und entspricht der Abweichung der stationär-interpolierten Projektion zur realen Emission.

Das Einsparpotenzial eines WLTC ist fast 10 Prozentpunkte geringer als das der betrachteten RDE-Ersatztests, da die dynamischen Anteile \dot{M} nicht so hoch sind wie bei einem RDE-Ersatztest (vergl. Abb. 3.20). Ein RDE-Ersatztest mit Anteilen hoher Lastgradienten kann nach dieser Analyse­methode ein NO_x -Einsparpotenzial von 37 % aufweisen. Der Einfluss des Fahrprogramms Sport oder Komfort beträgt 3 Prozentpunkte.

Ein entscheidender Faktor bei der Abschätzung der Potenziale ist die korrekte Zeitsynchronisation des NO_x -Signals um die Gaslaufzeit, bzw. Totzeit. Die Berechnung folgt einer Totzeitkorrektur mit der in Kap. 3.3 vorgestellten Parameterschätzung mit ARX-Modellen der zweiten Ordnung. Die Zahlen beziehen sich auf eine ideale, verlustfreie Beschreibung, sollte es eine Kalibrierung geben, die sämtliche Abweichungen dynamisch auffangen kann.

3.4.5 Bewertung der Bandbreite der Sensorik

Der stationäre Kalibrierprozess sieht Mittelwertmessungen für die Abgaskomponenten im konstanten Betriebspunktbetrieb vor. Die Messtechnik ist für statische Anwendungen ausgelegt. Für die Optimierung transienter Fahrmanöver ist die Eignung der Messtechnik auf die Emissionskomponente NO_x in dynamischer Anwendung zu prüfen. Die verfügbaren Messtechniken sind zum einen die Abgasmessanlage (AMA), die dem Funktionsprinzip eines *Chemilumineszenz Detektors* (CLD) folgt. Des Weiteren sind die PEMS mit dem Messprinzip der nichtdispersiven Infrarot-Spektroskopie (engl. *Non-dispersive ultra violet spectroscopy* NDUV) verfügbar sowie eine schnelle Abgasmesseinheit mit dem *Fourier transformations infrarotspektrometer* (FTIR). Zusätzlich ist der NO_x -Sensor untersucht. Das Wirkprinzip des NO_x -Sensors basiert auf einem Doppelkammeraufbau. In der ersten Kammer erfolgt die Trennung des Sauerstoffes vom Abgas durch eine Pumpspannung an der Hauptelektrode. Die abgepumpten Sauerstoffmoleküle durchdringen einen Zirkonoxid-Keramik-Festelektrolyt. In der hinteren Kammer wird das restliche NO und NO_2 in die Bestandteile Stickstoff und Sauerstoff gespalten. Der Sauerstoff wird ionisiert und durch eine heiße Messelektrode gepumpt. Der abgepumpte Massenstrom ist proportional zur NO_x -Konzentration [99]. Beim Vergleich der Messtechniken spielen deren Totzeit T_d sowie deren Anstiegszeit bzw. Ansprechverhalten T_{10-90} eine Rolle. Die Zeitkonstanten sind in Abb. 3.25 schematisch für ein PT1-Glied dargestellt. Die Anstiegszeit T_{10-90} entspricht der Zeitänderung von T_{10} bis T_{90} und sollte in Dynamikuntersuchungen möglichst gering sein. Die zur Verfügung stehenden Messeinheiten sind mit einigen Eigenschaften in Tab. 3.7 zusammengefasst.

Die Messgeräte detektieren Konzentrationen c_i , welche in einen Emissionsmassenstrom \dot{m}_i umgerechnet werden. Dafür werden die Signale um die Totzeit korrigiert.

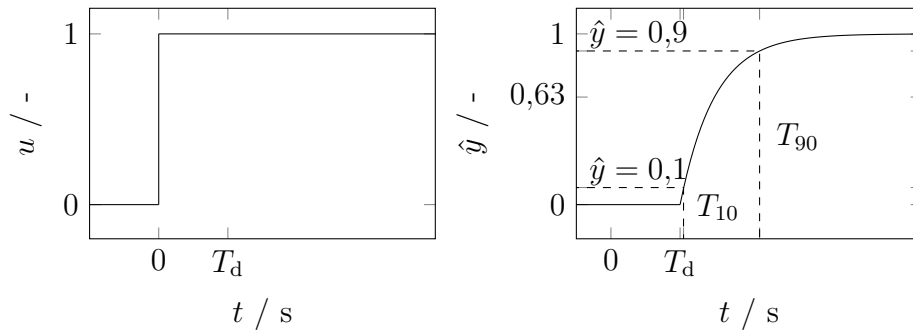


Abbildung 3.25 Sprungantwort eines PT1-Gliedes \hat{y} mit der Totzeit T_d sowie der Anstiegszeit T_{10-90} auf den Eingangs-Einheitssprung u

Tabelle 3.7 Verfügbare Messtechniken zur Betrachtung von NO_x -Emissionen.

	AMA	PEMS	FTIR	NO_x -Sensor
NO_x -Messung	NO/NO_x	NO, NO_2	$\text{NO}, \text{NO}_2, \text{N}_2\text{O}$	NO, NO_2
Messprinzipien	CLD	NDUV	FTIR	Messelektrode
Anstiegszeit T_{10-90}	1,8 s	1 – 3 s	0,2 s	0,5 s
Totzeit T_d	8 – 10 s	5 – 7 s	1 – 2 s	0,5 – 2 s
Anwendung	Standard	RDE	Sondermessung	Abgasanlage

Der Emissionsmassenstrom folgt der Beziehung der Stoffkonzentration und des Abgasmassenstrom mit der Beziehung für NO_x :

$$\dot{m}_{\text{NO}_x} = c_{\text{NO}_x} \dot{m}_{\text{exh}} \frac{\rho_{\text{NO}_2}}{\rho_{\text{exh}}} = c_{\text{NO}_x} \dot{m}_{\text{exh}} \frac{2,053 \text{ g dm}^{-3}}{1,2928 \text{ g dm}^{-3}} \quad (3.30)$$

Die Dichten ρ , bezogen auf die Temperatur 271 K und den Druck 1013 mbar, werden als Konstanten verwendet und sind aus der Motorsteuererätafunktion übernommen. Ein Vergleich der Sprungantworten von verschiedenen Messgeräten im Betriebspunkt $n = 1500 \text{ min}^{-1}$ und $M = 163 \text{ Nm}$ ist in Abb. 3.26 dargestellt. Die Sprünge werden mit einer Änderung der Gesamteinspritzmenge $q_{I, \text{Ges}}$ ausgelöst. Abgebildet sind die korrigierten NO_x -Massenströme für AMA, FTIR und den NO_x -Sensor, roh-seitig entnommen hinter der Abgasturbine. Die Messeinheiten geben einen ähnlichen Verlauf der Massenströme wieder. Der dargestellte Ausschnitt einer einzelnen Einspritzsequenz lässt die Unterschiede in den Maxima und im Ansprechverhalten erkennen. Die schnellste Reaktionszeit liefert das Messgerät FTIR. Die AMA zeigt träges Verhalten. Der NO_x -Seriensensor liegt zwischen den beiden Kurven. Speziell im Bereich des Sprungs weist der NO_x -Sensor im Dynamikverhalten Vorteile gegenüber der AMA auf. Rohemissionsseitig bietet die FTIR ein hohes Potenzial für die Dynamikuntersuchungen.

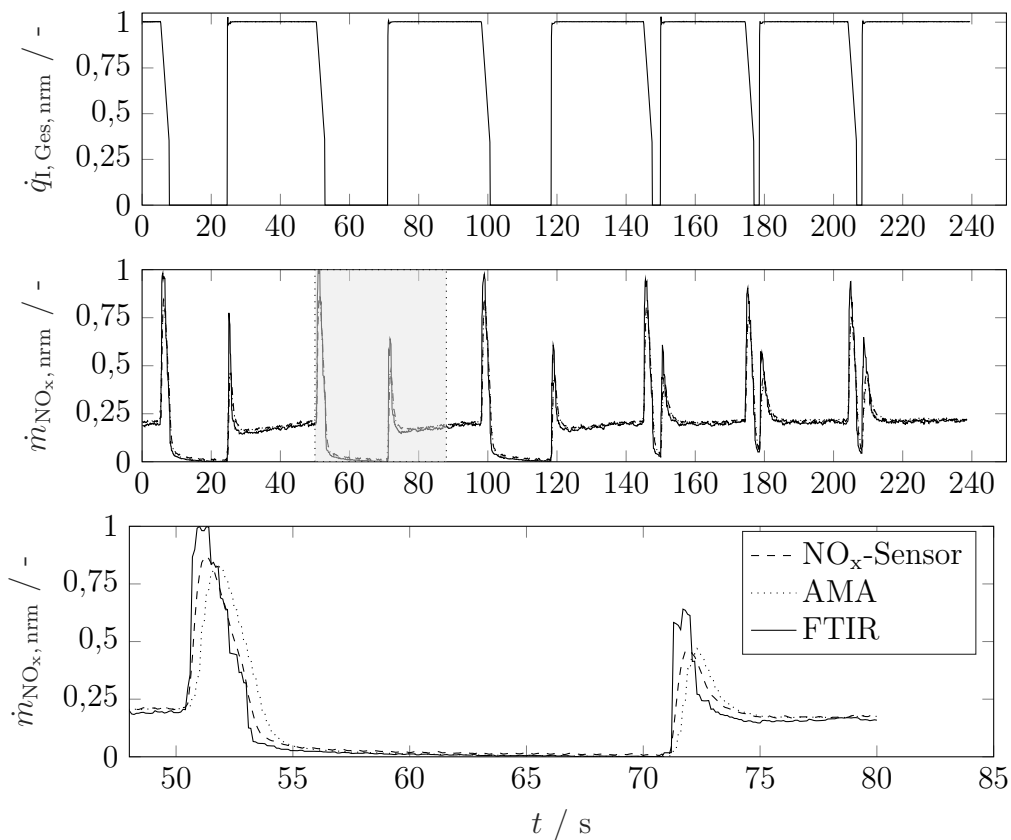


Abbildung 3.26 Vergleich der Rohemissionen des NO_x -Massenstroms gemessen mit dem NO_x -Sensor, AMA und FTIR auf die Sprünge der Gesamteinspritzmenge $q_{I, \text{Ges}}$

Heinz u. a. [50] beschäftigt sich mit dem Dynamikverhalten einer schnellen Sondermesstechnik und vergleicht zu dieser Referenz den NO_x -Sensor. Bei der schnellen Sondermesstechnik handelt es sich um den *Fast NO_x -Analyzer CLD500* der Firma *Cambustion* mit einer dynamischen Anspruchszeit $t_{10-90} = 2$ ms und einer Datenerfassung von bis zu 1 kHz. Die Forschungsarbeit klassifiziert den NO_x -Sensor ebenfalls als ausreichend für die Analyse dynamischer Fahrzyklen, wenn die schnelle Sondermesstechnik als wahre Referenz herangezogen wird [50].

3.5 Zusammenfassung der methodischen Untersuchung

Folgende methodische Fragestellungen wurden in diesem Kapitel beantwortet:

- Hohe Eingangsraumdimensionen von modellbasierten Kalibrieraufgaben können mit den Methoden der Datenanalyse sowie der Sensitivitätsanalyse reduziert werden, indem die Einflussparameter auf die wesentlichen begrenzt werden. Wird die Korrelationsanalyse verwendet, ist es ratsam, diese mit der verallgemeinerten Transinformation zu kombinieren, um nichtlineare statistische Zusammenhänge nicht zu vernachlässigen.
- Varianzbasierte Sensitivitätsmaße sind auf statische Modelle anwendbar und sollten im Fall von Versuchsraumgrenzen durch den Stichproben-Hyperkubus zur Berechnung der Modellvarianzen überprüft werden. Eine visuellen Überprüfung der Einhaltung der Versuchsraumgrenzen schaffen die beiden grafischen Erklärbarkeitsansätze PDP und ICE. Gegenüber den varianzbasierten Sensitivitätsanalysen bieten diese den Vorteil, die Richtung des Einflusses zu zeigen.
- Soll eine transiente Kalibrierung optimiert oder ein dynamisches Modell gebildet werden, sind Totzeiten zu berücksichtigen. Diese lassen sich stabil und in numerisch vertretbarem Rahmen mittels ARX-Modellen zuverlässig schätzen. Ist der Zusammenhang der Daten stark nichtlinear, muss die Schätzung mit numerisch aufwendigeren Methoden, z. B. der verallgemeinerten Transinformation, überprüft werden.
- Die Bedeutung der Dynamik in nichtsynthetisch erstellten RDE-Ersatztests wurde aufgezeigt. Der Einfluss der Dynamik bewirkt eine Abweichung von den stationär optimierten Betriebspunkten. Eine vorgeschlagene Erweiterung des Führungsgrößen-Definitionsraums um eine dynamische Größe hat das Potenzial, die entstehenden Emissionen zu reduzieren. Das vorgeschlagene dreidimensionale Führungsgrößen-Wertetripel bildet sich aus Drehzahl, Last und Lastgradient.
- Für eine modellbasierte Optimierung des transienten Fahrverhaltens wurde die verfügbare Messtechnik auf die Problemstellung NO_x verglichen und bewertet. Die Messdaten bescheinigen dem NO_x -Sensor eine hinreichende Bandbreite für dynamische Untersuchungen.

4 Optimierung von Emissionen im transienten Fahrbetrieb

In Kap. 3 ist der Betriebspunkt in Abhängigkeit von den Führungsgrößen Drehzahl und Lastgröße einer herkömmlichen Motorsteuergerätefunktion vorgestellt. Die bilineare Interpolation der Kalibrierparameter erfolgt abhängig von den beiden Führungsgrößen und unabhängig von der Dynamik. Im Folgenden soll die Erweiterung des Betriebspunktes zum Wertetriple Drehzahl, Last und Lastgradient am Prüfstand experimentell mittels einer modellbasierten Kalibrierung untersucht und optimiert werden. Dafür werden zunächst die Kalibrierparameter sowie die Dynamik des Prüflaufs festgelegt und mittels der Sensitivitätsanalyse auf einen Haupteinflussfaktor beschränkt. Da bei der transienten Kalibrierung mehrere Stützstellen eines Kalibrierparameters im Drehzahl-Last-Kennfeld entlang der Fahrspur variiert werden müssen, erhöht jede Stützstelle die Freiheitsgrade bzw. die Dimension des Versuchsaufwands. Anstelle einer Kennfeldgröße wird infolge der dynamischen Führungsgröße Lastgradient eine Kennraum-Lösung für den optimierten Kalibrierparameter vorgeschlagen. Im weiteren Verlauf werden die Versuchsumfänge an einem 2,0-Liter-Reihenvierzylinder Dieselmotor (OM 654 DE 20) mit einer zweistufigen Turboaufladung vorgestellt. Die Zielgröße der Versuchsreihe liegt in der Emissionsreduzierung des Abgasbestandteils NO_x und der anderen Emissionskomponenten über die transiente Fahrspur.

4.1 Auswahl der Kalibrierparameter und Versuchsplanung

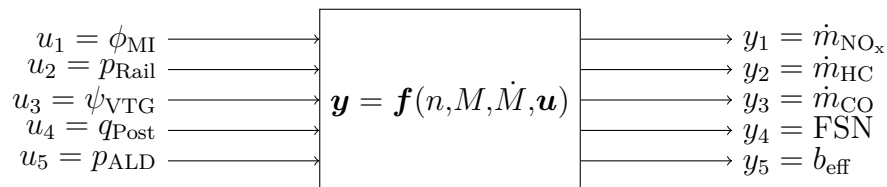


Abbildung 4.1 Möglicher Versuchsumfang der Kalibrierparameter und Zielgrößen für eine Kalibrierungsaufgabe der Grundmotorenbedatung

In Abb. 4.1 ist eine Kalibrierungsaufgabe für einen Dieselmotor mit fünf Kalibrierparametern \mathbf{u} sowie fünf Zielgrößen \mathbf{y} gezeigt. Im Eingangsraum werden die Kalibrierparameter EZMI ϕ_{MI} , der Raildruck p_{Rail} , die VTG-Rate ψ_{VTG} , die Einspritzmenge der Nacheinspritzung q_{post} und der Abgasladedruck p_{ALD} variiert. Die Zielgrößen sind die Massenströme der Emissionsbestandteile CO, HC, NO_x , der effektive spezifische Kraftstoffverbrauch b_{eff} sowie die Schwärzungszahl (engl. *Filter smoke number*, FSN). Die Schwärzungszahl entspricht der optoelektronischen Auswertung von geschwärztem Filterpapier über ein Reflexfotometer [97, S. 200]. Die Anzahl der Kalibrierparametern erhöht sich zudem durch die Führungsgrößen der Kennlinien und Kennfelder um die Drehzahl n und die Last M . Dieser Umfang wird durch die Berücksichtigung des Lastgradienten \dot{M} um eine weitere Größe erhöht. Anstelle stationärer Mittelwertmessungen bei einer stationären Kalibrierungsaufgabe werden transiente Fahrspuren zur Anregung der Dynamik verwendet. Dadurch steigert sich der zeitliche Aufwand des Versuchs verglichen mit der herkömmlichen Weise durch die Länge der Fahrspur sowie die Notwendigkeit, ein breites Spektrum an Drehzahl- und Lastgradienten in der Optimierung zu berücksichtigen [3, S. 114]. Der nachfolgende Versuchsaufwand wird daher auf positive Lastgradienten begrenzt. Die Höhe des Lastgradienten wird aus dem Profil eines RDE-Ersatzzyklus selektiert. Der Fokus liegt dabei auf positiven Laständerungen. Zudem wird mithilfe der Sensitivitätsanalyse die Problemstellung auf einen wesentlichen Kalibrierparameter bzw. Haupteinflussfaktor begrenzt. Hierbei erfolgt nachfolgend die Auswahl des Kalibrierparameters an den vorhandenen Rohdaten stationärer Prüfstandmessungen.

4.1.1 Sensitivitätsanalyse der Kalibrierparameter

In den Anwendungsbeispielen der Datenanalyse (Kap. 3.1.3) sowie der Sensitivitätsanalyse (Kap. 3.2.3) sind für stationäre Prüfstandsdaten die Kalibrierparameter bereits auf die Zielgröße NO_x untersucht. Die Daten stammen aus einer stationären modellbasierten Kalibrierung am selben Versuchsträger im Betriebspunkt $n = 1600 \text{ min}^{-1}$ und $M = 300 \text{ Nm}$. Beide Methoden identifizieren den EZMI ϕ_{MI} für den NO_x -Massenstrom im Variationsbereich des stationären Drehzahl-Last-Betriebspunkts mit dem höchsten Einfluss aller Kalibrierparameter. Die Ergebnisse der gesamten Sensitivitätsanalyse auf sämtliche in Abb. 4.1 aufgeführten Zielgrößen bzw. Modellausgänge sind in Tab. 4.1 dargestellt.

Die Sensitivitätsanalyse identifiziert den EZMI ϕ_{MI} für sämtliche Messgrößen \hat{y}_1 bis \hat{y}_5 als den größten Einfluss. Auf den NO_x -Massenstrom hat die VTG-Stellgröße ψ_{VTG} den zweitgrößten Einfluss. Gemeinsam mit dem EZMI ϕ_{MI} addiert sich der Einfluss der VTG-Stellgröße ψ_{VTG} auf eine Summe der beiden Haupteinflussfaktoren von $\sum S_i = 0,907$. Die Sollgröße der Abgasladedruckregelung p_{ALD} hat auf die Modellausgänge $\hat{y}_1 = \text{FSN}$ und den NO_x -Massenstrom geringe Einwirkung. Die Einspritzmenge der Nacheinspritzung q_{post} hat ebenfalls geringe Auswirkungen auf die Emissionsbestandteile.

Tabelle 4.1 Sensitivitätsanalyse für die Haupteinflüsse S_i der modellbasierten Kalibrierung mit den Kalibrierparametern u_i auf fünf Modellausgänge \hat{y}_i

Ausgänge	$u_1 = \phi_{\text{MI}}$	$u_2 = p_{\text{Rail}}$	$u_3 = \psi_{\text{VTG}}$	$u_4 = q_{\text{Post}}$	$u_5 = p_{\text{ALD}}$	$\sum S_i$
$\hat{y}_1 = \text{FSN}$	0,405	0,202	0,177	0,031	0,016	0,830
$\hat{y}_2 = \dot{m}_{\text{NO}_x}$	0,500	0,017	0,407	0,025	0,035	0,985
$\hat{y}_3 = \dot{m}_{\text{HC}}$	0,716	0,005	0,039	0,020	0,079	0,858
$\hat{y}_4 = \dot{m}_{\text{CO}}$	0,533	0,007	0,101	0,064	0,144	0,849
$\hat{y}_5 = b_{\text{eff}}$	0,538	0,011	0,014	0,037	0,392	0,992

Der höchste Einfluss ist auf den Modellausgang $\hat{y}_4 = \dot{m}_{\text{CO}}$ festzustellen mit $S_4 = 0,064$. Modelle mit schwachen Wechselwirkungen geben die Haupteinflüsse für die Modellausgänge $\hat{y}_2 = \dot{m}_{\text{NO}_x}$ und $\hat{y}_5 = b_{\text{eff}}$ wieder, da die Summe sämtlicher Indizes $\sum S_i \approx 1$ ergibt. In den restlichen Modellausgängen sind $1 - \sum S_i$ Interaktionseffekte zwischen den Kalibrierparametern festzustellen. Für ein besseres stationäres Systemverständnis sind die visuellen Erklärbarkeitsansätze PDP und ICE in Abb. 4.2 für die Modellausgänge \dot{m}_{HC} , \dot{m}_{CO} , FSN und b_{eff} gezeigt. Diese Kurvenscharen ermöglichen ein visuelles Verständnis des Zielkonfliktes der Optimierung.

Eine NO_x -Minderung kann mit einer Einstellung des Kalibrierparameters EZMI kleiner $\phi_{\text{MI}} < -6$ °KW erreicht werden. Dies bewirkt nach den in Abb. 4.2 gezeigten Daten gleichzeitig eine Erhöhung des Kraftstoffverbrauchs b_{eff} sowie eine Erhöhung der HC-Emissionen. Für $-6 < \phi_{\text{MI}} < 5$ °KW ist die rote Linie des PDP der HC-Emissionen unwesentlich beeinflusst und auf ähnlichem Niveau nahezu konstant. Die FSN-Zahl steigt bis zum EZMI von $\phi_{\text{MI}} \approx -3$ °KW und sinkt mit einer weiteren zeitlichen Verlagerung nach hinten. Der CO-Massenstrom lässt sich mit einem Polynom dritten Grades approximieren, mit einem Minimum im Scheitelpunkt zwischen $\phi_{\text{MI}} \approx -8$ und -5 °KW. In Kap. 3.2.3 ist der Effekt dieses Parameters ϕ_{MI} auf NO_x bereits in der ICE-Abb. 3.11a dargestellt. Eine zeitliche Verzögerung des EZMI ϕ_{MI} nach hinten reduziert den NO_x -Massenstrom. Aufgrund der Ergebnisse der Sensitivitätsanalyse wird der EZMI ϕ_{MI} als Kalibrierparameter der transienten Kalibrierung ausgewählt.

4.1.2 Auswahl der Lastgradienten

In Abb. 4.3a ist eine Clusteranalyse von Anteilen mit positiven Laständerungen eines RDE-Ersatzzyklus aufgezeigt. Der RDE-Ersatzzyklus wird in Segmente mit positiven Laständerungen unterteilt. Die Laständerungen sind mit der Ausgangsparametrierung vermessen und über der Drehzahl n und dem Drehmoment M sowie der Dauer t der Laständerung in t -Richtung abgebildet.

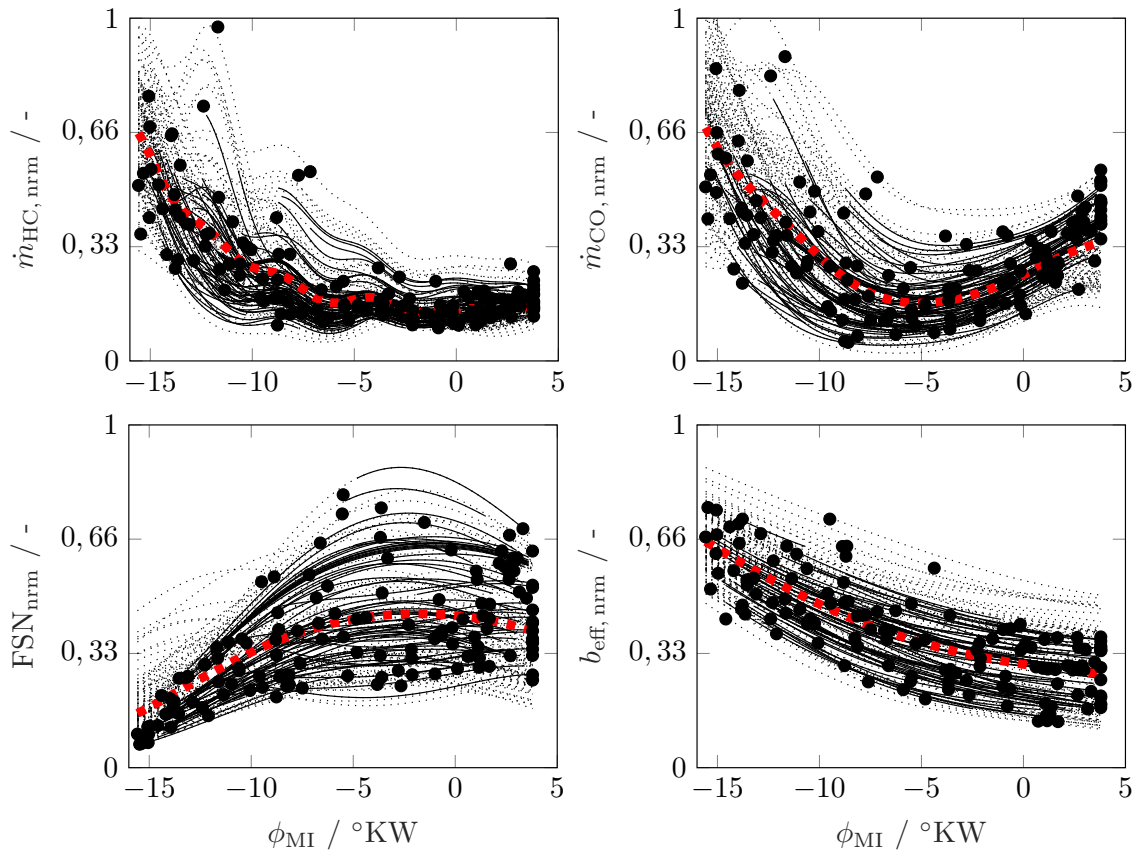


Abbildung 4.2 PDP (rot) und ICE-Plot (schwarz) für den EZMI ϕ_{MI} und die Auswirkungen auf die statischen Modellausgänge \dot{m}_{HC} , \dot{m}_{CO} , FSN und b_{eff} im Variationsbereich für eine stationäre Versuchsplanung ($\dot{M} = 0 \text{ Nm s}^{-1}$). Versuchsraumverletzungen sind gepunktet, gültige Modellauswertungen als Linie dargestellt

Dargestellt sind sämtliche Anteile von Laständerungen mit einer Mindestdauer von $t = 1 \text{ s}$ einer RDE-Ersatzzyklus-Fahrspur. In der Ebene $t = 0 \text{ s}$ liegen die Betriebspunkte im Drehzahl-Last-Kennfeld, aus denen die Anteile mit einer positiven Laständerung starten. Die Linien der Abb. 4.3a verbinden diese Ausgangspunkte mit den Betriebspunktwerten am Ende des Lastsprungs über der Dauer t der Laständerung. Die maximale Dauer ist $t = 12,5 \text{ s}$. Die Laständerungen reichen bis zu einer maximalen Last von $M = 410 \text{ Nm}$. Wird die Laständerung mit der Dauer t dividiert, entspricht dies dem Lastgradienten \dot{M} des RDE-Ersatzzyklus, gezeigt in Abb. 4.3b.

Der Lastgradient \dot{M} ist über der gemittelten Drehzahl der positiven Laständerungen in Abb. 4.3b dargestellt. Die kumulierte Höhe der NO_x -Emission ist normiert zu den jeweiligen Beschleunigungsanteilen farblich gekennzeichnet. Häufungen der Datenpunkte in vertikaler Richtung sind für die Start-Stopp-Anteile bei der Start-Drehzahl $n = 0 \text{ min}^{-1}$ sowie bei positiven Beschleunigungsvorgängen aus einer Leerlaufdrehzahl $n = 750 \text{ min}^{-1}$ zu erkennen.

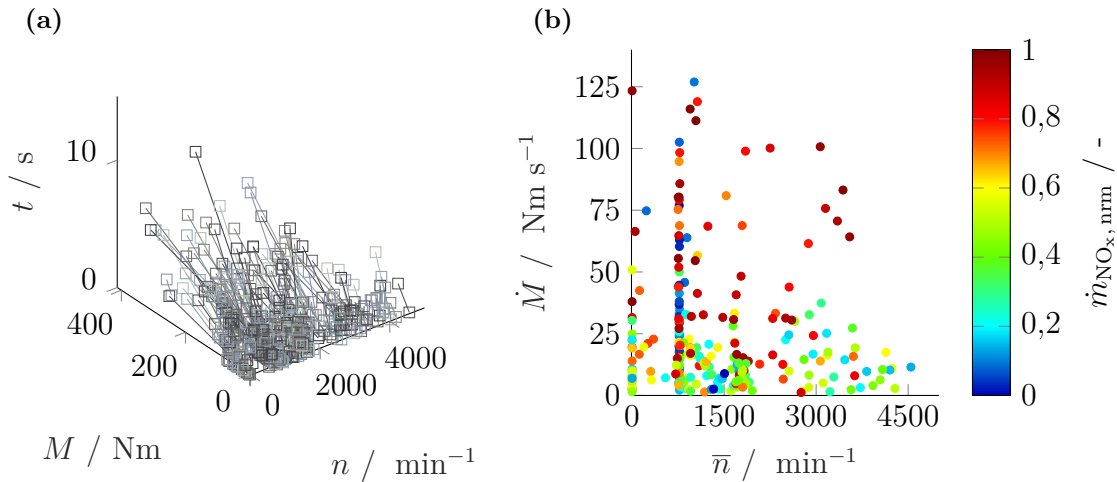


Abbildung 4.3 Analyse positiver Laständerungen der Anteile eines RDE-Ersatzzyklus im Drehzahl-Last-Kennfeld $n - M$ mit der Beschleunigungsdauer t (a) und deren zugehöriger Lastgradient mit kumulierter NO_x -Emission des Anteils über der gemittelten Drehzahl \bar{n} des Lastsprungs (b)

Der Lastgradient hat hier Werte zwischen $\dot{M} = 0 \text{ Nm s}^{-1}$ und $\dot{M} = 130 \text{ Nm s}^{-1}$. Eine weitere Häufung tritt in den Drehzahlen $n = 1400 \text{ min}^{-1}$ bis $n = 2000 \text{ min}^{-1}$ für Lastgradienten $\dot{M} < 30 \text{ Nm s}^{-1}$ auf. Einen Einfluss der Lastsprünge aus der Leerlaufdrehzahl $n = 750 \text{ min}^{-1}$ auf die Höhe der NO_x -Emission ist nicht festzustellen. Niedrige normierte NO_x -Emissionen $\dot{m}_{\text{NO}_x, \text{nrm}} < 0,5$ lassen sich für die Beschleunigungsvorgänge unter dem Lastgradienten $\dot{M} < 40 \text{ Nm s}^{-1}$ erkennen. Über dem Lastgradienten $\dot{M} > 40 \text{ Nm s}^{-1}$ sind die normierten NO_x -Emissionen $\dot{m}_{\text{NO}_x, \text{nrm}} > 0,6$ abgesehen von einigen Punkten im Bereich der Leerlaufdrehzahl. Eine Kalibrierung, welche den Lastgradienten als Führungsgröße berücksichtigt, verspricht eine Emissionseinsparung für verschieden hohe Dynamiken definiert über den Lastgradienten.

Anhand dieser Ergebnisse werden folgende Betriebspunkte bzw. Parameter der Laständerung des Motors für die Prüfstanduntersuchung festgelegt: Die Drehzahl ist konstant bei $n = 1600 \text{ min}^{-1}$. Die Laständerung, die den Lastgradienten bildet, erfolgt oberhalb einer Nulllast auf das Drehmoment von $M = 350 \text{ Nm}$. Abb. 4.4 zeigt die Istwertkurven des Drehmoments M der transienten Fahrspur für die konstante Drehzahl $n = 1600 \text{ min}^{-1}$ und die positive Laständerung über $M = 350 \text{ Nm}$. Die Dauer des Beschleunigungsvorganges auf die maximale Last wird in der Versuchsplanung mit $t = 5 \text{ s}$, $t = 3 \text{ s}$ und $t = 1 \text{ s}$ vermessen. Folglich umfasst der Versuchsaufwand die positiven Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$, $\dot{M} = 117 \text{ Nm s}^{-1}$ und $\dot{M} = 350 \text{ Nm s}^{-1}$. Die Kurve des Lastgradienten $\dot{M} = 350 \text{ Nm s}^{-1}$ ist im oberen Lastbereich $M > 350 \text{ Nm}$ abnehmend, da die Dynamik der Laständerung bereits von der Rauchbegrenzung des Motors abgeschwächt wird.

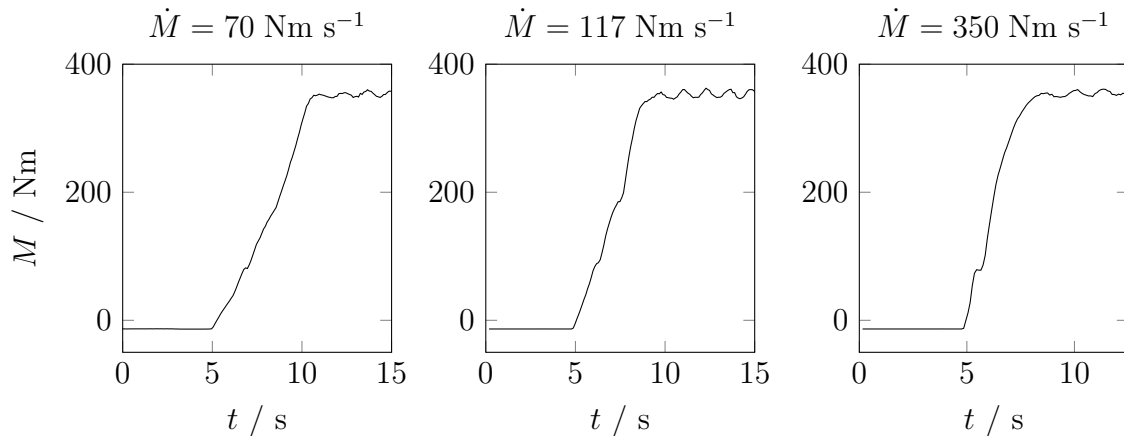


Abbildung 4.4 Vergleich der Istwertkurven des Drehmoments M der Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ bis $\dot{M} = 350 \text{ Nm s}^{-1}$ bei einer konstanten Drehzahl $n = 1600 \text{ min}^{-1}$

4.1.3 Versuchsplanung des Kalibrierparameters

In einer stationären modellbasierten Kalibrierung des EZMIs ϕ_{MI} wird ein Betriebspunkt eingestellt und der Wert für den Einspritzparameter in diesem Betriebspunkt variiert. Im transienten Fahrbetrieb ändert sich der Betriebspunkt über die Zeit und den Lastgradienten. Das Kennfeld des EZMI wird im dynamischen Betrieb durchfahren, und somit verändern sich die Sollwerte entsprechend der Form des Kennfeldes durch die bilineare Interpolation des Kalibrierparameters. Damit sind in der Versuchsplanung anstelle eines skalaren Wertes mehrere Kennfeldstützstellenwerte entlang der transienten Fahrspur in der Versuchsplanung zu berücksichtigen. Der Kalibrierparameter wird in Abhängigkeit vom Sollwert der Last über die Laständerung entlang der Sollwertkurve im Kennfeld variiert. Dies erhöht im Vergleich zum stationären Versuch die Freiheitsgrade des Versuchsaufwands für jede zu variiierende Stützstelle pro Kalibrierparameter.

Im Steuergerät wird der Kalibrierparameter EZMI ϕ_{MI} in einem Kennfeld definiert. Das Kennfeld ist über den beiden Führungsgrößen Drehzahl n und Einspritzmenge q_{MI} (Lastgröße) aufgespannt. Da der Versuchsumfang auf die konstante Drehzahl $n = 1600 \text{ min}^{-1}$ beschränkt wird, entspricht der Kalibrierparameter einem Schnitt des Kennfeldes bei dieser Drehzahl. Der Kalibrierparameter repräsentiert somit eine über der Einspritzmenge q_{MI} aufgetragene Kennlinie. Der transiente Anteil des Versuchs wird über die Laständerung auf $M = 350 \text{ Nm}$ realisiert. Abhängig vom Istwert der Lastgröße Einspritzmenge q_{MI} wird bei einer konstanten Drehzahl n der Sollwert des EZMIs ϕ_{MI} im Steuergerät aus der gedachten Kennlinie interpoliert. Die Beschleunigungsdauer der positiven Laständerung t gibt die Höhe des Lastgradienten vor. Für die drei Laständerungen der Abb. 4.4 sind die interpolierten Sollwertvorgaben des EZMIs q_{MI} in Abb. 4.5 dargestellt.

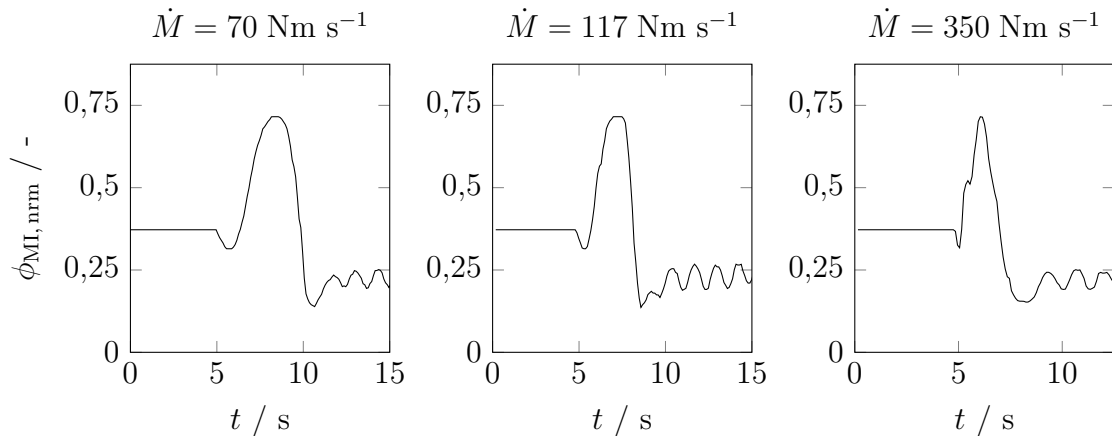


Abbildung 4.5 Vergleich der interpolierten Sollwertvorgabe des EZMIs ϕ_{MI} aus dem Kennfeld in Abhängigkeit der Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ bis $\dot{M} = 350 \text{ Nm s}^{-1}$

Die drei Kurven in Abb. 4.5 zeigen abhängig von der Höhe des Lastgradienten die bei diesem dynamischen Fahrmanöver aus dem Drehzahl-Last-Kennfeld interpolierten Sollwertkurven für den EZMI ϕ_{MI} . Die Kennfeldstützstellen entlang der transienten Fahrspur geben die Anzahl der Freiheitsgrade in der Versuchsplanung vor. Die Dimension der Versuchsplanung entspricht der Anzahl an Stützstellen der Motorsteuererüaterfunktion des jeweiligen Parameters. Bei der transienten modellbasierten Kalibrierung wird jeder Stützstellenwert einzeln variiert und unterschiedliche Kombinationen von Stützstellenwerten werden am Prüfstand vermessen. Da die Anzahl an Stützstellen die Dimension und damit die Anzahl an Parametern in der Versuchsplanung erhöht, steigt der Versuchsaufwand für transiente Untersuchungen selbst mit wenigen Kalibrierparametern erheblich. Deshalb wird im Folgenden eine Reduzierung der zu variiierenden Stützstellen des Eingangsparameters bzw. der Kennlinie vorgeschlagen.

In Abb. 4.6 sind für den Kennfeldschnitt bei der Drehzahl $n = 1600 \text{ min}^{-1}$ mehrere Kennlinien vom Kalibrierparameter EZMI ϕ_{MI} aus der Versuchsplanung dargestellt. Die Kennlinien sind über der Lastgröße Einspritzmenge q_{MI} jeweils normiert aufgetragen. Der Nullpunkt der y-Achse $\phi_{\text{MI}} = 0$ bezieht sich auf einen Wert hinter dem oberen Totpunkt des Kolbens. Der maximale Wert $\phi_{\text{MI}} = 1$ ist gleichbedeutend mit einer zeitlichen Verlagerung des EZMI nach vorne. Die Laständerung des transienten Prüflaufs startet in einem Wert zwischen der Einspritzmenge $q_{\text{MI}} = 0$ und $q_{\text{MI}} = 0,1$ und steigt bis zum Drehmoment $M = 350 \text{ Nm}$, der einem Wert der Einspritzmenge von $q_{\text{MI}} \approx 0,7$ entspricht. Im Wertebereich $q_{\text{MI}} > 0,6$ der Kennlinie werden die Stützstellen des EZMI nicht variiert und entsprechen exakt der Ursprungsparametrierung. Der Variationsbereich der Stützstellen in der Kennlinie für den EZMI ϕ_{MI} umfasst ausschließlich den Lastbereich der Laständerung, der dem grauen Bereich in Abb. 4.6 entspricht. Der Variationsbereich ist in der Versuchsplanung auf einen Limit im zeitlichen Kurbelwinkelbereich nach hinten beschränkt.

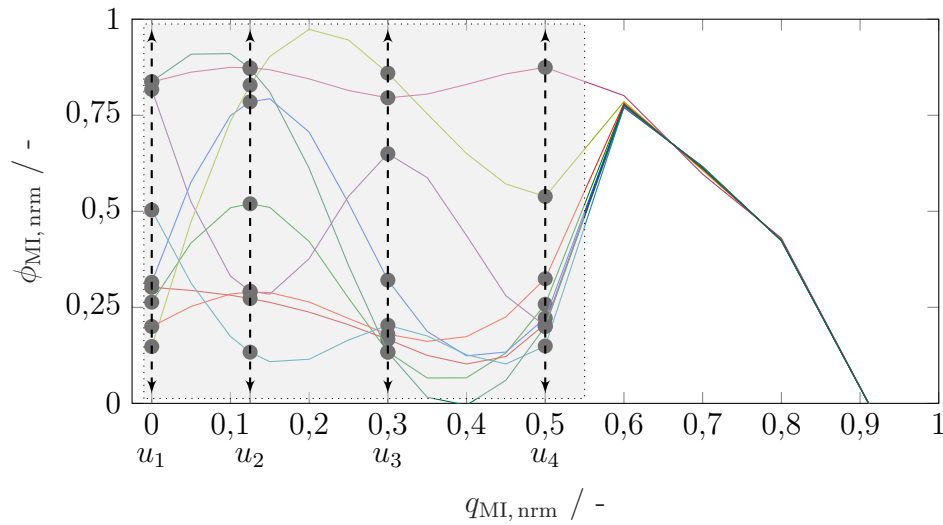


Abbildung 4.6 Ausgewählte, geglättete Kennlinien der Versuchsplanung für die Optimierung des EZMIs ϕ_{MI} , aufgetragen über der Einspritzmenge q_{MI} bei konstanter Drehzahl n

Mit der Beschränkung werden dynamische Verbrennungsschwankungen im Lastsprung unterbunden und der Versuchsträger geschont. Entlang der Kennlinie sind vier Stützstellen als Eingangsparameter $q_{MI,1} = u_1$ bis $q_{MI,4} = u_4$ bestimmt. Diese werden in einem statistischen Versuchsplan im normierten Variationsbereich $[0, 1]$ per *Latin hypercube sampling* variiert. Da kein Vorwissen über die zu parametrierenden Modellstruktur und das Prozessverhalten existiert, empfiehlt sich ein raumfüllender Versuchsplan gegenüber einem optimalen Design [69, S. 47, 48, S. 97]. Anstatt sämtliche Stützstellen zu variieren, reduziert sich die Versuchsplanung auf die vier Freiheitsgrade. Aus den vier variierten Stützstellenwerten werden mittels des in Kap. 2.3.3 vorgestellten Glättungsalgorithmus der gesamte Verlauf der Kennlinie und damit alle Stützstellenwerte berechnet. Die vier Stützstellenwerte entsprechen den Modelleingängen. Der Algorithmus behandelt die variierten Parameter u_1 bis u_4 sowie die Ursprungsparametrierung $q_{MI} > 0,6$ als exakt vorgegebene Punkte \mathbf{v}_{ex} der Randbedingung (vgl. Gl. 2.31).

In Abb. 4.6 sind mehrere mit der Methode der Kennfeldberechnung auf Basis der vier variierten Stützstellenwerte $q_{MI,1} = u_1$ bis $q_{MI,4} = u_4$ ermittelten Kennlinienverläufe gezeigt. Aus dem Kap. 3.4.2 sind die skalare Kennlinienverschiebung, die skalare Multiplikation und die Drei-Punkte-Kennlinienerstellung bekannt. Diese skalare Multiplikation entspricht einer Streckung oder Stauchung der Kennlinie. Die vorgeschlagene Strategie zur Erstellung von Kennlinien mit vier variierten Stützstellenwerten mit dem Glättungsalgorithmus erstellt universelle Kennlinienformen. Sie ist also nicht wie bei der Drei-Punkte-Kennlinienerstellung auf lineare oder quadratische Formen beschränkt. Der gesamte Messumfang beträgt 280 Variationen, 14 Wiederhol- und 24 Validierungspunkte, aus denen Kennlinien erstellt werden.

4.2 Modellierung, Optimierung und Validierung

Die jeweiligen Konzentrationen der Abgasmassenkomponenten CO_2 , CO und HC werden mit einer AMA aufgenommen und mit dem Abgasmassenstrom nach der Gl. 3.30 auf Emissionsmassenströme für die Modellausgänge transformiert. Die Parameterschätzmethode mit ARX-Modellen zweiter Ordnung wird zur Zeitverschiebung der Messsignale und Korrektur der Totzeit verwendet. Die Konzentration von NO_x wird mit dem NO_x -Sensor aufgezeichnet. Die FSN-Messtechnik ist geeignet, eine transiente Messung aufzuzeichnen, und wird daher nicht behandelt. Die Laufruhe des Versuchs wird über die Standardabweichung des indizierten Mitteldrucks überprüft, um instationäre Verbrennungsphänomene zu identifizieren. Die 280 verschiedenen Variationen des Kennlinienverlaufs bewirken nur geringe Änderungen im Verlauf des Emissionsmassenstroms von CO_2 . Die Variation der Kennlinie hat somit auf diesen Ausgangsparameter keinen Einfluss. Aus diesem Grund wird auf deren Modellierung bzw. Optimierung verzichtet.

Die Kurven der Wiederholpunkte für die Modellausgangsgrößen der Emissionsmassenströme \dot{m}_i sind für die Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ bis $\dot{M} = 350 \text{ Nm s}^{-1}$ in Abb. 4.7 dargestellt. Im Falle der CO -Emissionen weichen die Kurven kaum voneinander ab und überlappen in bestimmten Bereichen. Eine Erhöhung des Lastgradienten erhöht das Maximum des gemessenen Emissionsmassenstroms \dot{m}_{CO} . Gleiches gilt für den Emissionsmassenstrom \dot{m}_{HC} . Die HC -Verläufe weisen eine breite Verteilung der Messsignale auf, die mit einem steigenden Lastgradient abnimmt. Für den NO_x - und den HC -Massenstrom können im maximalen Lastbereich $M = 350 \text{ Nm}$ Schwankungen festgestellt werden, die auf die in Abb. 4.4 zu erkennende Schwingung des Abgasmassenstroms und des Drehmoments zurückzuführen sind.

Abb. 4.8 zeigt die Kurvenschar von zehn Kombinationen der vier Variationen des EZMIs ϕ_{MI} über der Laständerungen des Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$. Die zugehörigen Emissionsmassenströme \dot{m}_i von NO_x , CO und HC sind über den Lastsprung für die Variationen dargestellt. Erkennbar ist, dass die HC -Massenströme eine breite Verteilung ähnlich den Wiederholpunkten aufweisen. Daher kann für die Emissionskomponente HC kein Effekt der Variation auf die Emissionen beobachtet werden. Für CO sind unterschiedliche Anstiegsverhalten sowie ein Einfluss auf die Maxima der zehn Variationen festzustellen. Auch die Kurven der NO_x -Massenströme \dot{m}_{NO_x} verlaufen nebeneinanderliegend mit unterschiedlicher Steigung. Das Anstiegsverhalten durch die Variation des EZMIs ϕ_{MI} beeinflusst eine Wölbung mit einem lokalen Minimum zwischen $t = 5 \text{ s}$ und 10 s . Eine Wölbung dieser Art ist in den Kurven des Drehmoments in Abb. 4.4 ebenfalls zu erkennen.

Die Bildung stationärer Modelle aus Messdaten reicht in der Regel für eine Vielzahl von Problemstellungen in der Kalibrierung von Motorsteuergeräten aus [59]. Für dynamische bzw. transiente Kalibrierprobleme besteht die Möglichkeit, individuelle Gütekriterien in Form von Skalaren zu bestimmen und diese in statische Kalibrierprobleme umzuwandeln [59].

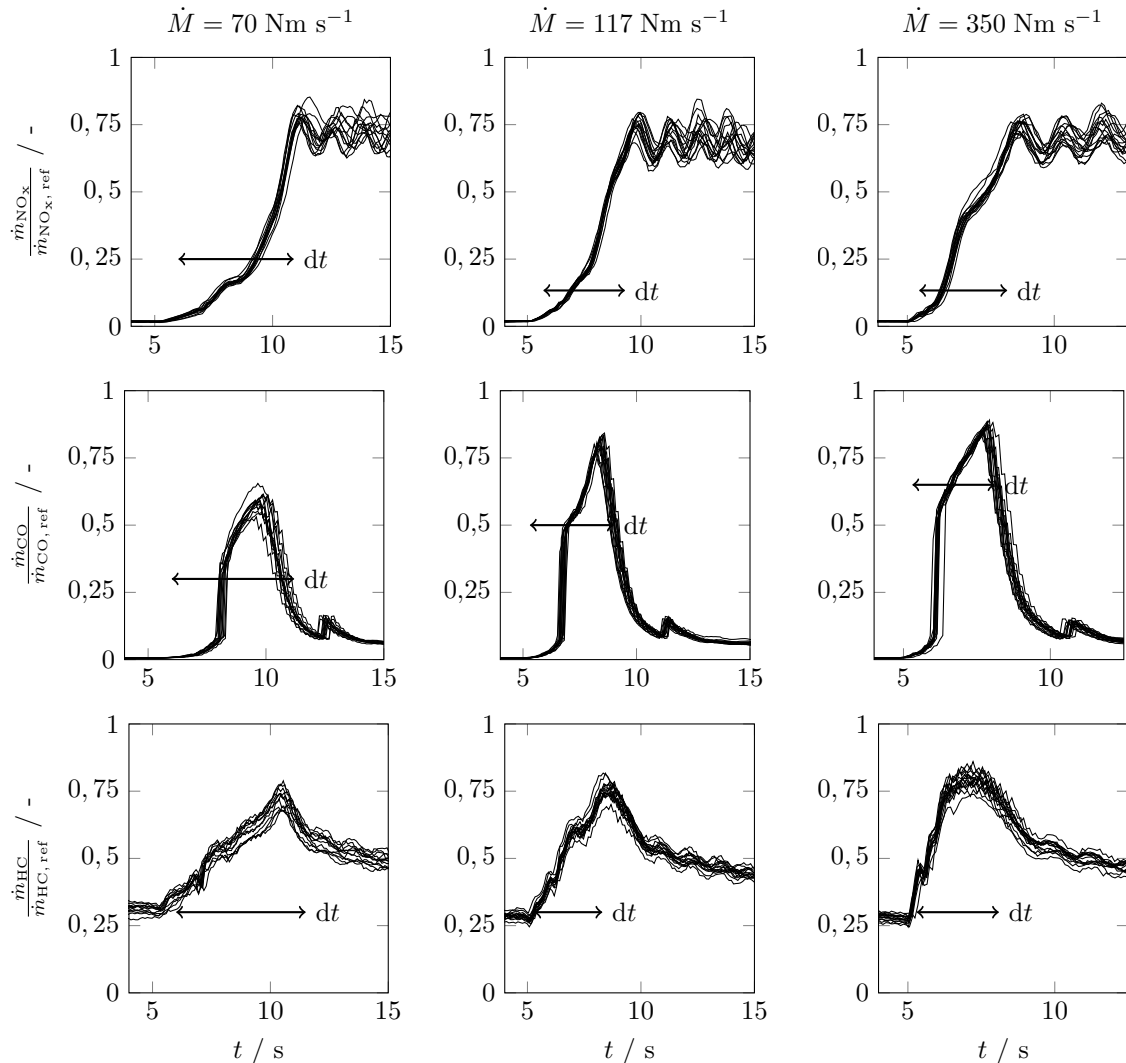


Abbildung 4.7 Vergleich der Emissionsmassenströme \dot{m}_i von NO_x , CO und HC abhängig vom Lastgradienten für die Wiederholpunkte

Statt einen Betriebspunkt zu vermessen und Mittelwerte von Kriterien als Modellausgang zu modellieren, wird ein transienter Verlauf gefahren und der Skalar beispielsweise durch einen Maximalwert ($\max(\dot{m}_i)$) oder Integralwert ($\int \dot{m}_i dt$) des Massenstroms \dot{m}_i gebildet. Die Reduzierung des transienten Verlaufs auf eine skalare Größe ermöglicht eine statische Modellierung anstelle einer deutlich komplexeren dynamischen.

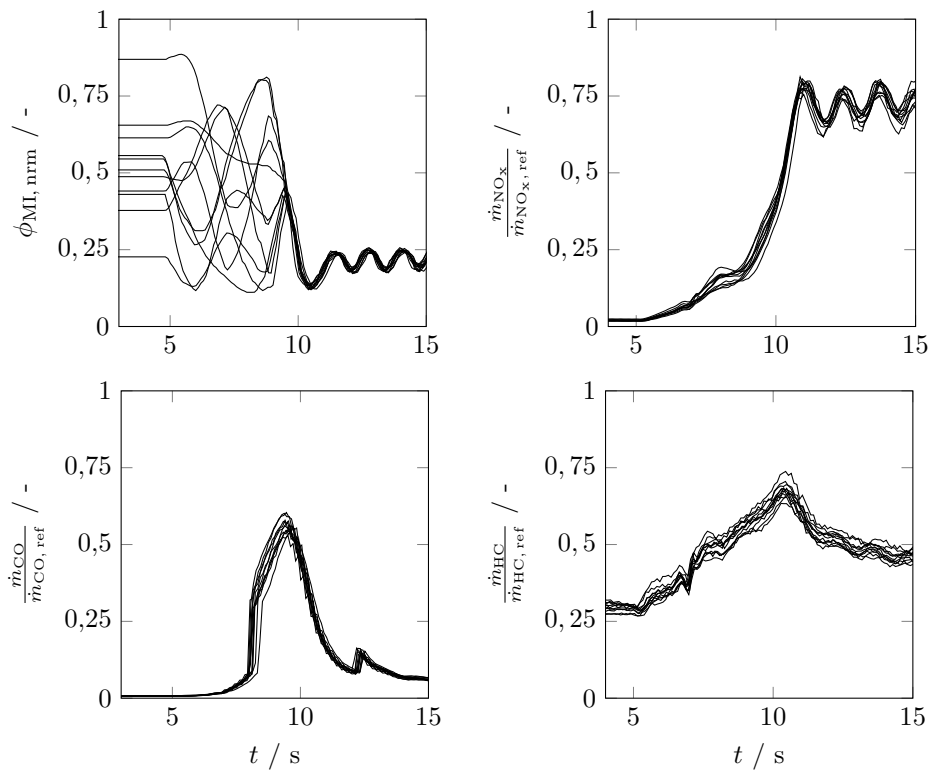


Abbildung 4.8 Einfluss von zehn Variationen des EZMIs ϕ_{MI} auf die Emissionsmassenströme \dot{m}_i von NO_x , CO und HC für den Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$

4.2.1 Statische Modellierung

Im Folgenden werden für die Emissionsmassenströme von NO_x , HC und CO die integralen Skalare $\int \dot{m}_i dt$ der Messdaten modelliert. Die Integrale werden vom Beginn der positiven Laständerung über das Intervall bis einschließlich zum Maximum der Massenstromkurven gebildet. Für jede Zielgröße wird an jedem Lastgradienten ein eigenes Modell gebildet. Da die Anstiegs- und Totzeiten der Messprinzipien unterschiedlich für jede Abgaskomponente sind, wird für jeden Emissionsbestandteil eine angepasste Integrallänge dt bis zum Maximum der Massenstromkurve festgelegt. Die Länge des Integrals dt ist für jede einzelne Zielgröße gleich lang (vgl. Tab. 4.2). Zur statischen Modellierung der skalaren Größen werden die in Kap. 2.1 vorgestellten Modellstrukturen Polynom, GPM und LMN verwendet und verglichen. Eine Übersicht der Modellfehler sowie der jeweils verwendete Modelltyp der Zielgröße ist in Tab. 4.2 gegeben. Als Modellfehler ist die normierte Wurzel der quadratischen Abweichung (engl. *Normalized root mean squared error* NRMSE) zwischen Modell- und Messwert gegeben. Zudem wird ein Kennwert zur Bewertung der Anpassung des Modells an einen externen, in der Modellschätzung nicht berücksichtigten Stichprobensatz (engl. *Predicted sum of squares* PRESS) herangezogen (siehe Kap. B.2 im Anhang) [106, S. 183].

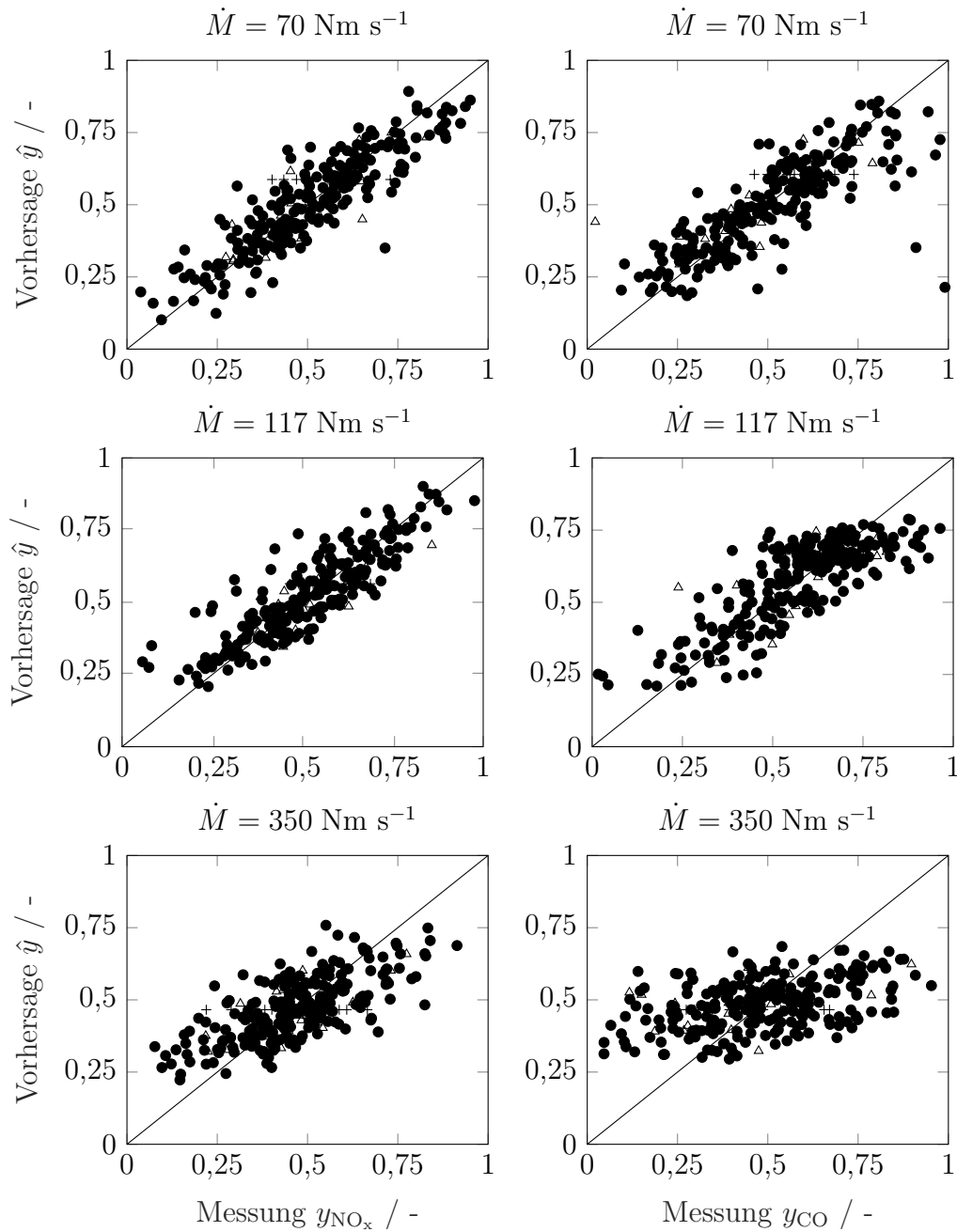


Abbildung 4.9 Streudiagramme der Modellprädiktionen \hat{y}_i über den Messwerten y_i der skalaren NO_x - (links) und CO -Größen (rechts) in Abhängigkeit vom Lastgradienten \dot{M} für Validierungsdaten (\triangle), Wiederholpunkte ($+$) und Messwerte (\bullet)

Streudiagramme der Modellprädiktionen \hat{y} über den Messwerten y der integralen NO_x - und CO -Größen sind in Abb. 4.9 gezeigt. Der Vertrauensbereich der jeweiligen Modelle ist in Tab. A.1 im Anhang angegeben. Die Abweichung der Punkte im Streudiagramm Abb. 4.9 sowie die Modellfehler der NO_x -Modelle zeigen für die Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ und $\dot{M} = 117 \text{ Nm s}^{-1}$ eine mittelmäßige Vorhersage

der Modelle. Die Gegenüberstellung der Mess- y und Modellwerte \hat{y} für den Modellfehler NRMSE weisen eine gering bis mäßige Streuung der Punkte bzw. Residuen um die 45°-Gerade auf. Anhand der Wiederholpunkte kann Prozessrauschen festgestellt werden. Für den Lastgradienten $\dot{M} = 350 \text{ Nm s}^{-1}$ werden niedrige Emissionswerte des NO_x -Modells überschätzt bzw. werden die hohen Emissionswerte unterschätzt. In diesem Fall ist der Lastgradient entweder zu hoch für die Dynamik der Messtechnik, oder der Einfluss der Variation des Kalibrierparameters auf die Emission ist zu gering. Von allen statischen Modellierungen weisen die NO_x -Modelle den geringsten Modellfehler NRMSE in Tab. 4.2 auf und sind in der Lage, das Prozessverhalten am besten abzubilden.

Tabelle 4.2 Prozentualer Modellfehler NRMSE der statischen Modellierung des transienten integralen Skalars der jeweiligen Zielgrößen \hat{y} für die Messdaten (FIT), Validierungsdaten (VAL) und für den Modellfehler PRESS, sowie Modelltyp und Integrallänge dt des modellierten Skalars

\hat{y}_i	dt	\dot{M}	Typ	$\text{NRMSE}_{\text{VAL}}$	$\text{NRMSE}_{\text{FIT}}$	PRESS
$\int \dot{m}_{\text{NO}_x} dt$	49 s	350 Nm s^{-1}	LMN	13,1	11,0	13,5
$\int \dot{m}_{\text{NO}_x} dt$	31 s	117 Nm s^{-1}	GPM	7,6	7,9	7,5
$\int \dot{m}_{\text{NO}_x} dt$	25 s	70 Nm s^{-1}	GPM	9,1	7,7	8,9
$\int \dot{m}_{\text{HC}} dt$	55 s	350 Nm s^{-1}	Polynom	18,3	18,3	18,8
$\int \dot{m}_{\text{HC}} dt$	40 s	117 Nm s^{-1}	Polynom	19,2	17,5	19,8
$\int \dot{m}_{\text{HC}} dt$	28 s	70 Nm s^{-1}	GPM	18,6	18,7	18,8
$\int \dot{m}_{\text{CO}} dt$	52 s	350 Nm s^{-1}	GPM	19,4	19,0	19,1
$\int \dot{m}_{\text{CO}} dt$	37 s	117 Nm s^{-1}	GPM	10,2	9,1	10,5
$\int \dot{m}_{\text{CO}} dt$	28 s	70 Nm s^{-1}	LMN	10,3	8,5	10,6

Das Modellverhalten der skalaren CO-Kriterien ist rechts der NO_x -Modelle in Abb. 4.9 gezeigt. Die Wertepaare der Modellprädiktion und der Messungen weisen für die Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ und $\dot{M} = 117 \text{ Nm s}^{-1}$ eine mittelmäßige Streuung um die 45°-Gerade auf. Der Modellfehler NRMSE ist höher als bei den NO_x -Modellen, siehe Tab. 4.2. Einige höhere Emissionswerte werden im Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ durch das Modell nicht erklärt. Im Lastgradienten $\dot{M} = 350 \text{ Nm s}^{-1}$ ist das CO-Modell nicht in der Lage, die Messwerte korrekt zu präzisieren.

Die Modellierung der integralen HC-Emissionen sagt die Messungen unzureichend vorher. Die Modellresiduen des Modellfehlers NRMSE sind in Abb. A.5 im Anhang aufgezeigt. Die Modellfehler sind in Tab. 4.2 angegeben. Die unzureichende Modellgüte kann entweder durch ein schlechtes Dynamikverhalten der AMA erklärt werden. Alternativ hat der gewählte Kalibrierparameter in transienten Fahrmanövern keinen entscheidenden Einfluss auf die Emission. Aufgrund der schlechten Modellgüte für den Lastgradienten $\dot{M} = 350 \text{ Nm s}^{-1}$ wird die Modellierung mit einem nichtlinearen dynamischen Ansatz überprüft.

4.2.2 Nichtlineare dynamische Modellierung

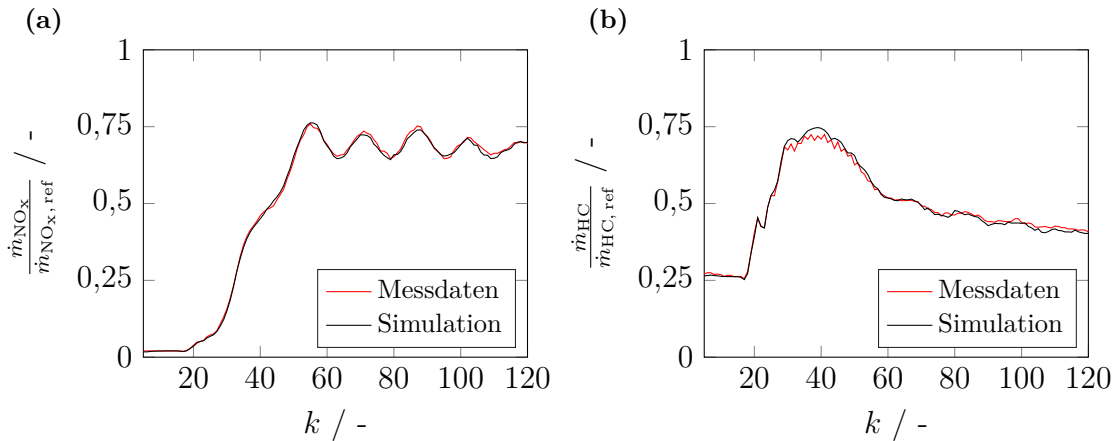


Abbildung 4.10 Gegenüberstellung der Emissionsmassenströme der transienten Prüfstandmessung und der Ausgänge der NFIR-Modellsimulation über dem diskreten Zeitschritt k für \dot{m}_{NO_x} (a) und \dot{m}_{HC} für einen Lastsprung

Infolge der ungenügenden Modellqualität der statischen Modellbildung der Emissionsgrößen für den Lastgradienten $\dot{M} = 350 \text{ Nm s}^{-1}$ werden die Signaldaten mit einer nichtlinearen dynamischen Modellstruktur approximiert. Anstelle der integralen skalaren Gütekriterien wird bei der dynamischen Modellbildung der abgetastete Signalverlauf mit einer NFIR, NOE oder NARX-Struktur abgebildet. Dafür wird eine *Matlab*-KNN-Umgebung von [42] verwendet. Die Umgebung erstellt eine Modellstruktur mit Hilfe von Lipschitz-Quotienten für das zu modellierende Prozessverhalten. Über die Lipschitz-Quotienten wird die dynamische Modellordnung bestimmt. Die Modelle sind einzig für den betrachteten Lastgradienten \dot{M} als lokale Lösung gültig. Die Anregung der Modelle ist ausschließlich für die lokale Lösung gültig und unzureichend für eine Beschreibung des kompletten Lastgradienten-Bereichs von $\dot{M} = 0 \text{ Nm s}^{-1}$ bis $\dot{M} = 350 \text{ Nm s}^{-1}$. Alle abgetasteten Datenpunkte werden für das Modelltraining verwendet, wobei jeder fünfte abgetastete Datenpunkt als Validierungspunkt zur Bewertung der Vorhersage der geschätzten Modelle herangezogen wird. Einige Variationsbereiche werden als Testdaten abschließend zur Überprüfung herangezogen und nicht im Modelltraining verwendet.

In Abb. 4.10 ist das nichtlineare dynamische Modell der Ausgangsgrößen der Emissionsmassenströme \dot{m}_{NO_x} (Abb. 4.10a) und \dot{m}_{HC} (Abb. 4.10b) für eine Testdatensequenz gezeigt. Die beiden Kurven zeigen eine Messung sowie das Ausgangssignal einer Modellsimulation. Die Eingangssignale der Modellbildung sind die Istwertkurven des Drehmomentes M und die Istwertkurven des EZMIs ϕ_{MI} . Die NFIR-Modellstruktur weist ein Bestimmtheitsmaß von $R_{\text{FIT}}^2 = 0,985$ auf. Die Validierungsdaten weisen einen geringeren Wert von $R_{\text{VAL}}^2 = 0,981$ auf. Die Wurzel der mittleren quadratischen Abweichung, engl. *Root mean square error*, (RMSE) ergibt für die

Messdaten $\text{RMSE}_{\text{FIT}} = 0,0004$ und für die Validierungsdaten $\text{RMSE}_{\text{VAL}} = 0,0005$. Die Modellgüte des Modells für NO_x ist in den abgebildeten Sprüngen der geschätzten sowie gemessenen Massenströme \dot{m}_{NO_x} und \dot{m}_{HC} zu erkennen. Eine geringfügige Abweichung zwischen der Simulation und der Messung ist für die HC-Signale \dot{m}_{HC} im Maximum zu beobachten. Das MLP-Netz des NFIR-Modells weist $n_M = 20$ Neuronen auf. Die Eingangsregressoren der TDNN-Filterbank sind mit Lipschitz-Quotienten auf die Anzahl $\varphi_i = 120$ festgelegt, siehe Kap. C.4 im Anhang. Die Parameter des Modells sind mit dem numerischen Optimierungsalgorithmus LMA trainiert.

4.2.3 Optimierung der empirischen Modelle

Ziel der Optimierung ist die Minimierung der Emissionen über den Lastgradienten \dot{M} gegenüber der Ursprungskalibrierung. Die Optimierung der statischen Modelle erfolgt mit dem genetischen multikriteriellen Optimierungsansatz (GMO). *IAV EasyDoE* ist über eine *Matlab*-Schnittstelle mit dem Optimierungsalgorithmus verbunden. Für die Modellausgänge NO_x , HC und CO ergibt sich für jeden Lastgradienten \dot{M} eine dreidimensionale Pareto-Front. Aus den Pareto-Fronten werden unterschiedliche Punkte entlang der Front verteilt ausgewählt, deren optimierte Eingangsparameter gewählt und am Prüfstand validiert. In Tab. 4.3 sind die Werte einer optimierten Parametrierung gezeigt. Die Parametrierungen minimieren sämtliche Kriterien und sind gleichzeitig maximal auf das integrale NO_x -Kriterium gewichtet ist. Die prozentualen Ergebnisse der Optimierung sind in Tab. A.1 im Anhang aufgeführt.

Die Optimierung des nichtlinearen dynamischen Modells für den Lastgradienten $\dot{M} = 350 \text{ Nm s}^{-1}$ erfolgt in der *Matlab*-KNN-Umgebung von [42] mittels der Partikelschwarmoptimierung (PSO). Da bei der Optimierung zwei Zielgrößen definiert sind, die gleichzeitig zu optimieren sind, werden diese auf eine einzige, übergeordnete Zielgröße reduziert. Eine Gewichtung der Zielfunktion des Optimierungsproblems für NO_x und HC erfolgt mit einem Verhältnis der Mittelwerte der Messreihen. Der iterative Suchpfad der PSO für die Zielfunktionsauswertung ist in Abb. A.4b im Anhang gezeigt. Die optimalen Lösungen der statischen sowie dynamischen Modelle sind in Tab. 4.3 für den GMO und die PSO aufgezeigt.

Tabelle 4.3 Validierte Parametrierung der Optimierungen der statischen und dynamischen Modellen des jeweiligen Lastgradienten sowie der Typ des numerischen Optimierungsalgorithmus

\dot{M}	Typ	u_1	u_2	u_3	u_4
70 Nm s ⁻¹	GMO	0,52	0,06	0,05	0,03
117 Nm s ⁻¹	GMO	0,51	0,05	0,04	0,03
350 Nm s ⁻¹	PSO	0,50	0,01	0,11	0,03

Die in Tab. 4.3 aufgeführten Optima zeigen eine Tendenz zur zeitlichen Verlagerung des EZMIs ϕ_{MI} nach hinten an den Rand der Versuchsraumgrenzen für die Stützstellen u_2 bis u_4 . Der auf den Variationsbereich normierte Wert der variierten Stützstelle u_1 liegt für die statischen Modellansätze $\dot{M} = 70 \text{ Nm s}^{-1}$ und $\dot{M} = 117 \text{ Nm s}^{-1}$ bzw. den nichtlinearen Modellansatz bei $\dot{M} = 350 \text{ Nm s}^{-1}$ und $u_1 = 0,5$. Berücksichtigt werden muss hierbei, dass der Lastsprung nicht aus der Nulllast $M = 0 \text{ Nm}$ erfolgt, sondern aus einem niedrigen Lastbereich. Der Startwert des Lastsprungs liegt somit zwischen den Stützstellen u_1 und u_2 . Die resultierenden optimierten Kennlinien, die mit dem Glättungsalgorithmus der Versuchsplanung erstellt werden, sind in Abb. 4.11 mit der Ursprungsparametrierung verglichen. Die optimierten Parametrierungen liegen für die drei Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ bis $\dot{M} = 350 \text{ Nm s}^{-1}$ deckungsgleich.

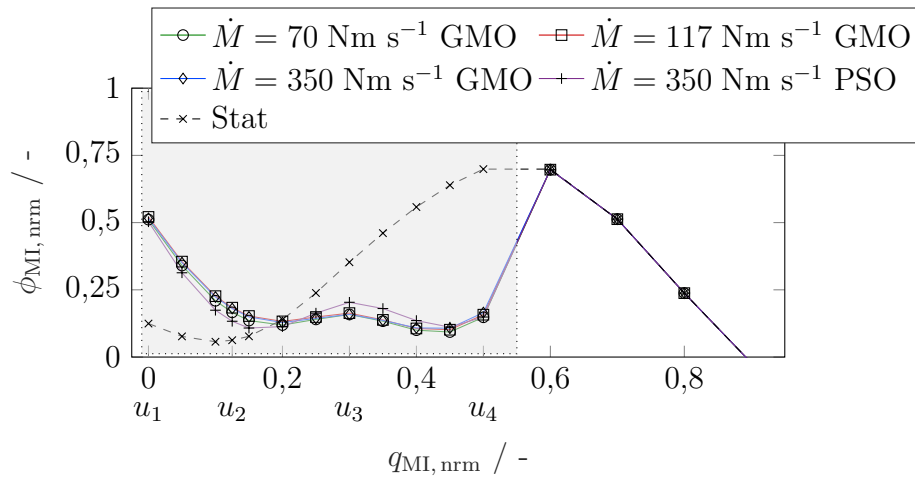


Abbildung 4.11 Die Ergebnisse der Kennfeldstützstellen im Variationsbereich für die optimierten Kriterien im transienten Fahrbetrieb in Abhängigkeit von der Dynamik des Lastgradienten \dot{M} im Vergleich zur stationär optimierten Ursprungsparametrierung (Stat)

4.2.4 Validierung und Interpretation der Ergebnisse

An die Optimierung schließt sich die Validierung der Parametrierung am Prüfstand an. Die aus den optimierten Stützstellen u_1 bis u_4 erstellten Kennlinien der Abb. 4.11 werden im Motorsteuergerät hinterlegt. Die Validierung vergleicht die Mittelwerte von jeweils vier integralen Skalaren der Emissionsmassenströme $\int \dot{m}_i dt$ der optimierten Kennlinien sowie von vier der Ursprungsparametrierung. In Abb. 4.12 werden die gemittelten Massenströme \dot{m}_i der optimierten Parametrierung mit der Ursprungsparametrierung verglichen. Die zugehörigen skalaren Integralwerte $\int \dot{m}_i dt$ der Validierungsmessungen sind in Abb. 4.13 gegenübergestellt. Die Integralwerte sind in Referenz zu den Skalaren der Ursprungsparametrierung gesetzt und skaliert angegeben.

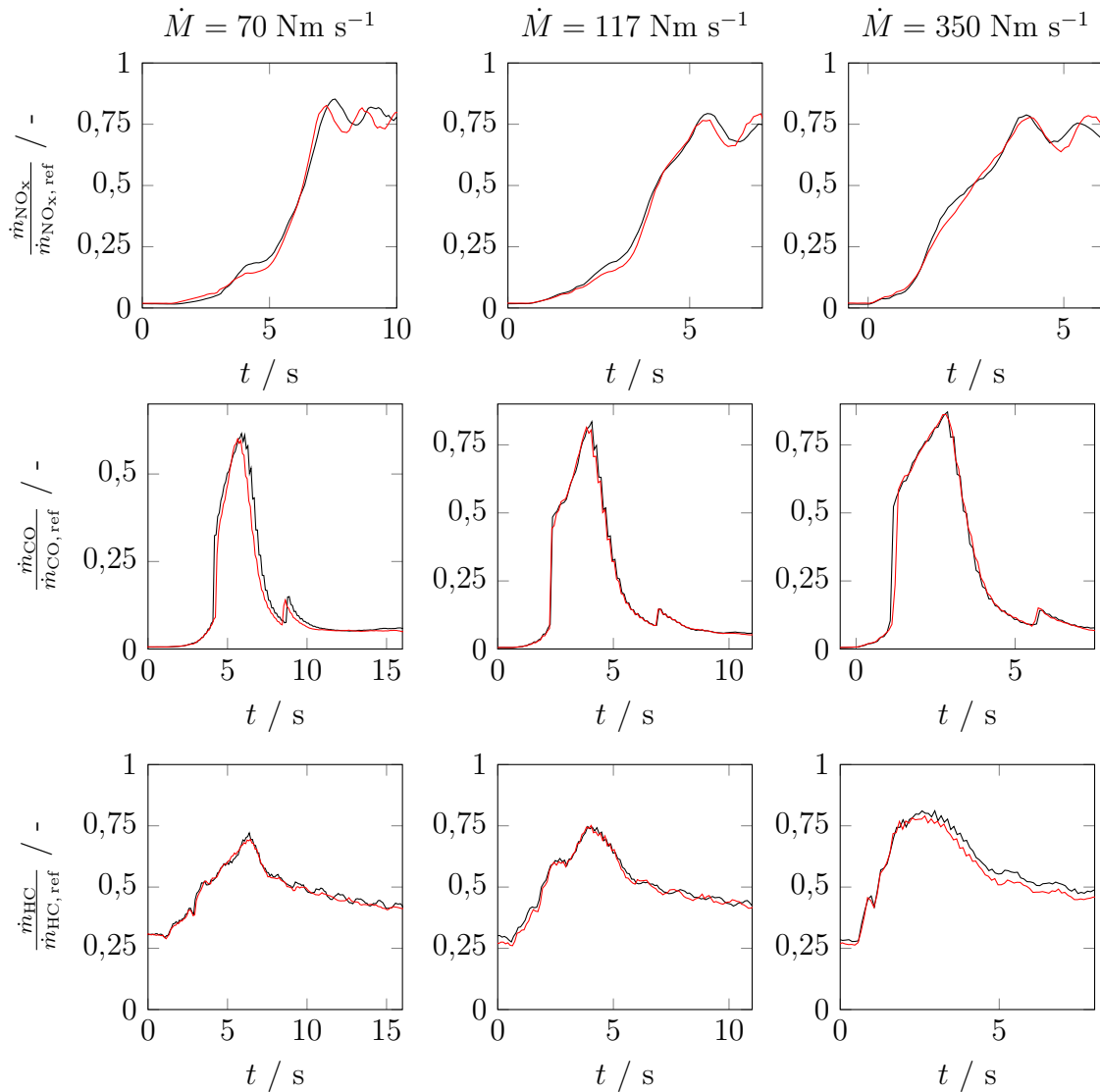


Abbildung 4.12 Vergleich von vier Mittelwertmessungen der stationär optimierten Ursprungsparametrierung (schwarz) mit der optimierten Parametrierung (rot) für die Emissionsmassenströme von NO_x , CO und HC in Abhängigkeit vom Lastgradienten \dot{M}

Für sämtliche Zielgrößen mit der optimierten Parametrierung für u_1 bis u_4 wird eine Verbesserung der integralen Skalare der NO_x -Emissionen erreicht, siehe Abb. 4.13d. Eine Kennfeldstrategie der zeitlichen Verzögerung des EZMI nach hinten im transienten Anspruchsverhalten reduziert die NO_x -Emissionen der skalierten Skalare im Bereich von 7 bis 9% für die Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ und $\dot{M} = 117 \text{ Nm s}^{-1}$. Die gemittelten Kurven der beiden Lastgradienten verlaufen unterhalb der Ursprungsparametrierung. Die optimierte Parametrierung reduziert die Wölbung im Anstieg. Zudem wird im Hochpunkt des gemittelten Emissionssprungs ein geringerer Wert erzielt.

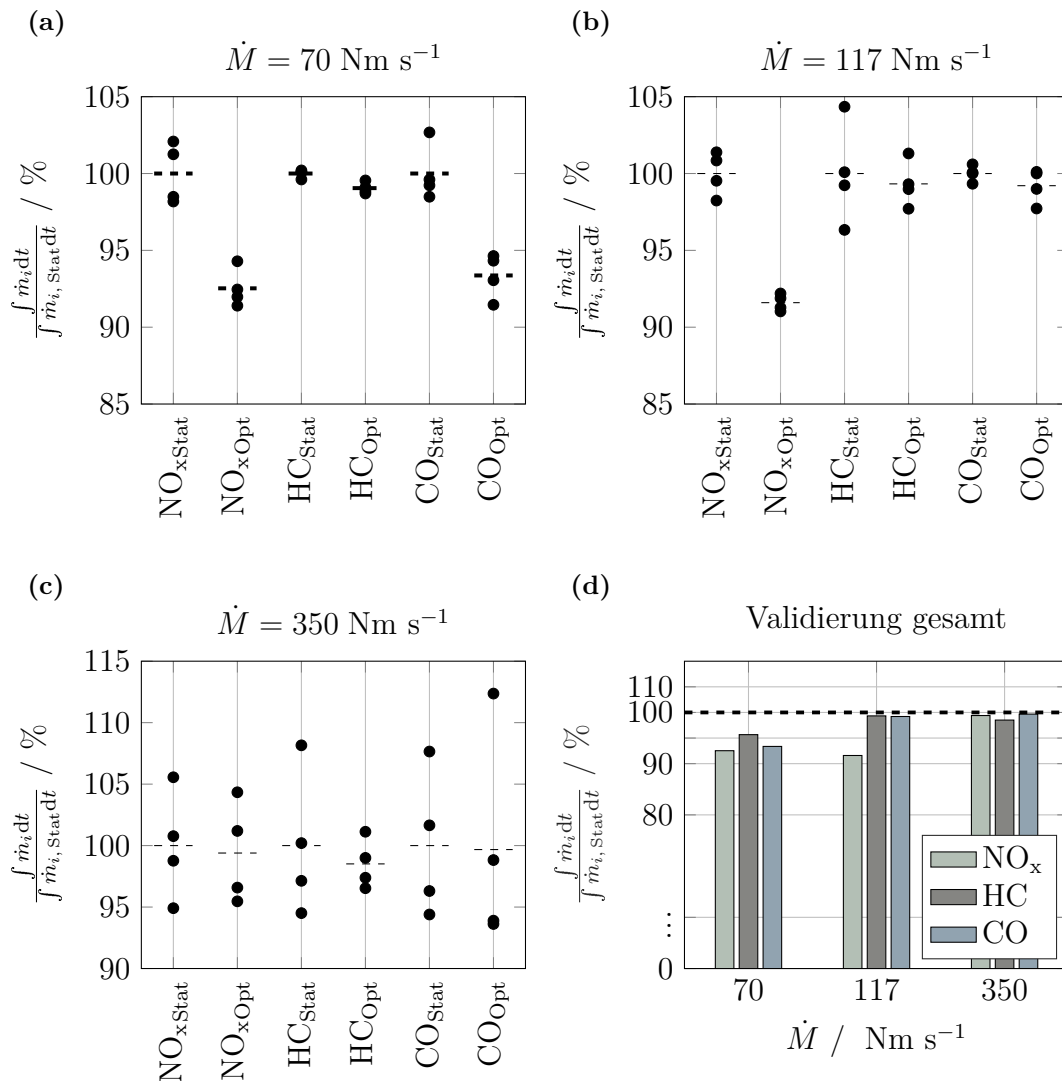


Abbildung 4.13 Skalierte Ergebnisse der Validierungsmessungen der Optimierungen (Opt) für die integralen NO_x -, HC- und CO-Kriterien im Vergleich zur stationär optimierten Ursprungsparametrierung (Stat) für die Lastsprünge $\dot{M} = 70 \text{ Nm s}^{-1}$ (a), $\dot{M} = 117 \text{ Nm s}^{-1}$ (b) und $\dot{M} = 350 \text{ Nm s}^{-1}$ (c), sowie zusammengefasst der Vergleich der Mittelwerte der jeweiligen Validierungsmessungen als Säulendiagramm (d)

Für den Lastgradienten $\dot{M} = 350 \text{ Nm s}^{-1}$ ist eine breite Streuung der integralen NO_x -Skalare in Abb. 4.13c um den Mittelwert festzustellen. Der gemittelte Wert der vier Messungen liegt 0,01% unter der Ursprungsparametrierung.

Eine Reduktion der CO-Kriterien ist für den Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ zu erkennen. Die Einsparung der optimierten integralen CO-Kriterien gegenüber der Ursprungsparametrierung liegt bei 6,7%. Für die höheren Lastgradienten $\dot{M} = 117 \text{ Nm s}^{-1}$ und $\dot{M} = 350 \text{ Nm s}^{-1}$ liegen optimierte und stationäre Ursprungsparametrierung nahezu deckungsgleich übereinander, mit ähnlichen Maxima der \dot{m}_{CO} -Kurven, vgl. Abb. 4.12, 4.13b und 4.13c. Dies ist zum einen auf die maximale Gewichtung der Optimierung auf die NO_x -Kriterien zurückzuführen. Zum anderen ist für den Lastgradienten $\dot{M} = 350 \text{ Nm s}^{-1}$ eine sehr breite Streuung der CO-Kriterien zu erkennen. Die beiden Parametrierungen weisen eine ähnlich breite Standardabweichung auf.

Die Mittelwerte der skalaren HC-Kriterien weisen eine Reduktion von 4,3% für den Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ auf. In den Abb. 4.13b und 4.13c ist eine breite Streuung der skalierten Integralwerte für sowohl die Ursprungsparametrierung als auch die optimale Parametrierung festzustellen. Die Streuung wirkt sich auf die in Abb. 4.12 dargestellten gemittelten \dot{m}_{HC} -Kurven aus. Die Kurven unterscheiden sich im Startwert des Lastsprungs bzw. für den Lastgradienten $\dot{M} = 350 \text{ Nm s}^{-1}$ im abfallenden Bereich nach dem Maximum. Dies ist auf die Messstreuung der gemessenen Signale zurückzuführen.

Zusammenfassend bestätigt die Validierung der optimierten Parametrierungen eine geringe Verbesserung der NO_x -Emissionen für die Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ und $\dot{M} = 117 \text{ Nm s}^{-1}$. Die HC- und CO-Kriterien bleiben dabei ähnlich der Ursprungsparametrierung. Im Lastgradienten $\dot{M} = 350 \text{ Nm s}^{-1}$ sind die Messstreuungen von der Ursprungsparametrierung und der optimierten Parametrierungen ähnlich breit gestreut. Die Messergebnisse sind stark verrauscht und lassen keine verlässliche Aussage bezüglich einer Optimierung zu.

4.3 Erweiterung des Betriebspunktes

Eine Erweiterung der Motorsteuererätefunktion von einer zweidimensionalen Kalibrierung im Drehzahl-Last-Kennfeld zu einer neuen Definition des Betriebspunktes als Wertetripel mit dem Lastgradienten als dynamische Führungsgröße ist vielversprechend. Die Ergebnisse der drei lokalen optimierten Parametrierungen sind nahezu gleich. Folglich ist in diesem Betriebspunkt die Parametrierung für positive Laständerungen unabhängig von der Höhe des Lastgradienten. Entsprechend gibt es im betrachteten Beispiel eine Lösung für verschiedene positive Lastgradienten \dot{M} .

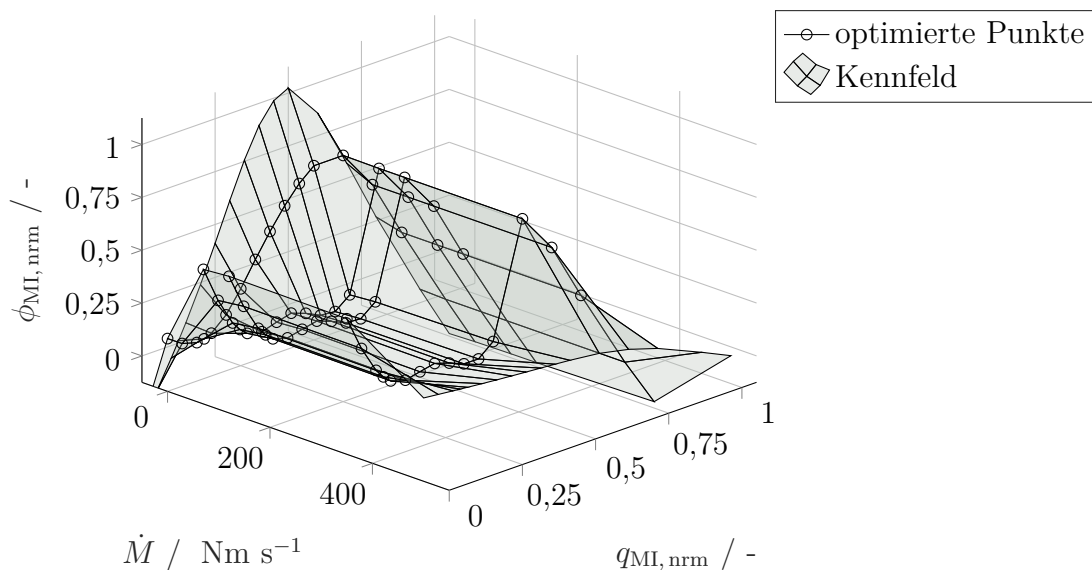


Abbildung 4.14 Projektion einer geglätteten Kennraumlösung der Erweiterung einer Motorsteuererätefunktion für den EZMI ϕ_{MI} zu einer Kalibrierung des Wertetripels unter Berücksichtigung des Lastgradienten \dot{M} bei einer konstanten Drehzahl $n = 1600 \text{ min}^{-1}$

Eine Projektion eines solchen Kennraums mit einer Dimension weniger kann exemplarisch die Abb. 4.14 visualisieren. Anstelle des Wertetripels ist hier die Drehzahl konstant gehalten, wodurch das Kennfeld der Abb. 4.14 die Lösung des Kennraums bei $n = 1600 \text{ min}^{-1}$ aufspannt. Eine Fortsetzung der Untersuchung auf den gesamten Drehzahlbereich sowie auf negative Lastgradienten ist folglich die Gesamtlösung einer kompletten Kennraumbeschreibung. Die Betrachtung von lokalen Betriebspunkten in Abhängigkeit von den Führungsgrößen Drehzahl, Last und Lastgradient ist in eine globale Betrachtung zu überführen, beispielsweise mit der Kombination von Anregungssignalen zur Systemanregung (Kap. 2.2.3). Die Versuchsplanung ist entsprechend auf diese dynamische Größe zu erweitern. Dies entspricht einer universellen Motorkalibrierung unabhängig vom realen Fahrzustand oder -profil und ist folglich für zukünftige RDE-Anforderungen geeignet.

5 Zusammenfassung und Ausblick

Im Mittelpunkt dieser Arbeit steht die Entwicklung einer Methode zur modellbasierten Optimierung von Antriebssystemen im transienten Betrieb. Mit einem geeigneten Ansatz sind im dynamischen Betrieb optimale Verbrauchs- und Emissionsgrößen realisierbar. Zur Erarbeitung der Zielsetzung wurde eine methodische Abschätzung der durch eine transiente Optimierung erreichbaren Emissionsreduktion vorgestellt. Mit dieser Potenzialabschätzung ist es möglich, die durch das Verlassen des quasistationär optimierten Drehzahl-Last-Betriebspunktes erhöhten Emissionen quantitativ zu bewerten. Das Einsparpotenzial einer Optimierung des transienten Betriebes ist abhängig von der Dynamik des betrachteten Zyklus. Für die Optimierung des dynamischen Betriebs auf Emissionen ist die entsprechende Sensorik essenziell. Daher wurde hier auch eine Bewertung der verfügbaren Messtechnik bezüglich der Messdynamik und Bandbreite im Hinblick auf die in RDE-Ersatzzyklen vorkommende maximale Momenten-Dynamik vorgenommen. Die Messdatenanalyse von Sprungtests und RDE-Sequenzen weist den NO_x -Sensor als hinreichend für Dynamikuntersuchungen am Prüfstand aus.

Die Applikationsaufgabe einer Optimierung von Emissionen im transienten Betrieb umfasst einen mehrdimensionalen, durch die Kalibrierparameter aufgespannten Eingangsraum. Daher folgt eine Untersuchung verschiedener Ansätze der (nicht-) linearen Korrelations- und Sensitivitätsanalyse zur Ermittlung der für die Emissionen relevanten Kalibrierparameter. Vorgestellt wurden methodische Ansätze, um den jeweiligen Einfluss von Parametern auf Ausgangsgrößen zu quantifizieren. Der informationstheoretische Ansatz der verallgemeinerten Transinformation bewertet statistische Abhängigkeiten in nichtlinearen Datensätzen. Zudem bietet die Transinformation ein Kriterium für die Ermittlung einer geeigneten Abtastung von Messdaten im *Downsampling*-Prozess der Datenvorverarbeitung. Die Schwächen der Gleichverteilung als Hyperkubus des Monte-Carlo-Stichprobenumfangs bei der Schätzung der varianzbasierten Sensitivitätsmaße sind grafisch mittels PDP und ICE kompensierbar. Aus Messdaten der Grundmotorenbedatung wird der EZMI als Haupteinflussfaktor auf die Diesel-Emissionskomponente NO_x identifiziert. Mit den Erklärbarkeitsansätzen kann die Einflussgüte des Haupteinflussfaktors auf den Verbrauch und die restlichen Emissionskomponenten visualisiert werden.

Die stationäre Optimierung von Motorsteuergeräten in den Drehzahl-Last-Arbeitspunkten geht von Mittelwertmessungen aus. Bei der Erweiterung zu einem Wertetripel mit Drehzahl, Last und Lastgradient als Führungsgrößen gilt es, Gaslaufzeiten zwischen den Emission-Massenströmen und den Motorgrößen bei transi-

enten Manövern zu berücksichtigen. Dazu wurde eine Untersuchung von Ansätzen zur Schätzung der Totzeit transienter Messungen vorgestellt. Die Parameterschätzverfahren mit ARX-Modellen zeichnen sich durch Effizienz, Stabilität sowie eine zweckdienliche Rechenzeit aus und sind folglich für transiente Betrachtungen heranzuziehen.

Ein transienter modellbasierter Applikationsansatz zur Optimierung eines Dieselmotors am Prüfstand wurde vorgestellt. Die Dynamik des Drehmoments ist einer der variierten Kalibrierparameter. Es wurden drei repräsentative Lastgradienten bei konstanter Drehzahl auf die Zielgrößen NO_x , HC und CO optimiert und deren Ergebnisse mit der stationären Basisapplikation verglichen. Dabei lassen sich mit den optimierten Parametern geringe Einsparungen der NO_x -Emissionen realisieren. Zum Einsatz kommen statische DoE-Methoden. Die Anwendung von nichtlinearen Modellansätzen resultiert in ähnlichen Optima, bietet jedoch keine Verbesserung im Vergleich zu den Ergebnissen der statischen Modellansätze.

Aus den Ergebnissen der in dieser Arbeit vorgestellten Optimierung von Emissionen im transienten Betrieb sowie in Zusammenschau mit den Veröffentlichungen [84, 105] bleibt festzuhalten, dass eine stationär optimierte Kalibrierung in einer dynamischen Anwendung weiteres Optimierungspotenzial bietet. Die optimierten Parametrierungen der drei verschiedenen Lastgradienten sind nahezu identisch. Die stationär optimierte Ursprungsparametrierung steigt im Vergleich zu den transient optimierten Parametrierungen ab der Stützstelle u_2 steil an und verschiebt zeitlich den EZMI nach hinten. Folglich kann von einer optimalen Lösung der transienten Kalibrieraufgabe unabhängig von dem Lastgradienten \dot{M} für die transiente Laständerung des betrachteten Bereiches bei der konstanten Drehzahl $n = 1600 \text{ min}^{-1}$ ausgegangen werden. Die Einsparung ist jedoch von geringer Bedeutung im betrachteten Beispiel. Erweitert werden kann der Versuchsaufwand durch weitere in Abb. 4.1 vorgestellte Variationsparameter sowie weitere Lastgradienten- und Drehzahlbereiche. Inwieweit eine Hinzunahme weiterer Kalibrierparameter zu einer höheren Einsparung führt, gilt es in einer zukünftigen Untersuchung zu prüfen.

In dieser Arbeit wurde eine Strategie der Kennlinien- und Kennfelderstellung mit einem Glättungsalgorithmus in der Versuchsplanung vorgeschlagen, mit der universelle Kennlinienformen erstellt werden können. Die Berechnung von Kennlinien ist somit nicht auf lineare oder quadratische Formen beschränkt und kann mit einem Glattheitsmaß kombiniert werden. Der Vergleich der optimierten Kennlinien des EZMIs mit der Ursprungsparametrierung ist in Abb. 4.11 zu sehen. Die Kennlinienform ist weder durch eine Stauchung oder Streckung, noch mit einer Verschiebung der Ursprungsparametrierung um einen skalaren Faktor zu erzielen. Die Variation der Stützstellenwerte in Kombination mit einem Glättungsalgorithmus verhindert steile Gradienten in den Kennlinien oder Kennfeldern.

Zudem bietet diese Variationsform ein geeignetes Verfahren, um unabhängig von der Ursprungsparametrierung eine optimierte Parametrierung zu finden. Die Versuchsplanung kann auf wenige Kennfeldstützstellen beschränkt werden, um verschiedene Formen von geglätteten Kennlinien oder -feldstrukturen experimentell zu untersuchen. Der Eingangsgrößenraum der Versuchsplanung ist somit beliebig skalierbar. Außerdem ist bei der Betrachtung der Lösung eine Reduzierung der Stützstellen-Variationsparameter u_1 bis u_4 auf die drei Parameter u_1 , u_3 und u_4 denkbar.

Die Grenzen der dynamischen Emissionsuntersuchung liegen im Dynamikverhalten bzw. im Ansprechverhalten T_{10-90} der Messtechnik und in der resultierenden Messstreuung. Mit steigendem Lastgradienten \dot{M} ist eine breitere Messstreuung der skalaren Kriterien festzustellen. Dies zeigen die Modellqualität in Tab. 4.2 und die Streudiagramme der Modellprädiktionen über den Messwerten in Abb. 4.9. Daher sollte der Versuch mit einer höheren Anzahl an Variationen erfolgen, um die Modellqualität zu erhöhen. Zudem ist es mit einer schnelleren Messtechnik in dynamischen Kalibrieraufgaben möglich, das Messrauschen zu reduzieren. Im Falle einer ungenügenden Qualität der statischen Modelle durch die integralen Skalare ist die Modellierung durch einen nichtlinearen dynamischen Ansatz in Kombination von TDNN-Filterbank und MLP-Netz erfolgt. Zwar steigt der numerische Rechenaufwand, hingegen entfällt der Aufwand der Berechnung geeigneter Skalare für die statische Modellierung. Der Lastgradient $\dot{M} = 350 \text{ Nm s}^{-1}$ ist entweder zu hoch für die Messtechnik oder der EZMI hat im hochdynamischen kaum oder gar keinen Einfluss auf die Emissionsmassenströme. Dadurch sind die Messdaten zu stark verrauscht, um eine Optimierung daraus ableiten zu können.

Aus den Ergebnissen der transienten, Lastgradienten-abhängigen Optimierung folgte der Vorschlag einer Erweiterung des Arbeitspunktes von Drehzahl und Last hin zum Wertetripel (Drehzahl/Last/Lastgradient). Eine abgeleitete Projektion eines solchen Kennraums wurde für eine konstante Drehzahl exemplarisch aufgezeigt. Eine Zyklusvorhersage kann mittels modellbasierter Kalibrierung mit dreidimensionaler Clusteranalyse eines Zyklus im Wertetripel-Kennraum erfolgen. Die Werkzeugkette der modellbasierten Applikation wurde entsprechend angepasst.

Für eine zukünftige Optimierung von Motorsteuergeräten im transienten Betrieb ist es entscheidend, die Abgassensorik für eine höhere Bandbreite weiterzuentwickeln. Gerade Ergebnisse zur schnellen NO_x -Abgasmesstechnik zeigen großes Potenzial für weitere Dynamikanwendungen und eine Minderung der Messstreuung [50]. Somit kann die Modellierung der Abgaskomponente verbessert werden. Sind Messeinrichtungen mit hoher Dynamik vorhanden, sollte eine Steuergeräte-Softwarestruktur für eine Wertetripel Strategie mittels z. B. eines *Rapid Control Prototyping*-Systems aufgebaut werden. Mit diesem System ist die Optimierung des gesamten Eingangsparameterraums auf noch weitere Zielgrößen durchführbar und eine Optimierung auf den gesamten transienten Betriebsbereich möglich. Eine weitere Senkung der Emissionen ist dabei für die Berücksichtigung weiterer Eingangsparameter in der Versuchsplanung zu erwarten, z. B. der VTG-Stellgröße oder mit einer schnellen AGR-Regelung.

Literatur

- [1] Amtsblatt der Europäischen Union. „Verordnung (EG) Nr. 715/2007 des Europäischen Parlaments und des Rates: vom 20. Juni 2007: über die Typgenehmigung von Kraftfahrzeugen hinsichtlich der Emissionen von leichten Personenkraftwagen und Nutzfahrzeugen (Euro 5 und Euro 6)“. In: Amt für Veröffentlichungen der Europäischen Union.
- [2] Amtsblatt der Europäischen Union. „Verordnung (EU) 2016/427 Der Kommission: vom 10. März 2016: zur Änderung der Verordnung (EG) Nr. 692/2008 hinsichtlich der Emissionen von leichten Personenkraftwagen und Nutzfahrzeugen (Euro 6)“. In: Amt für Veröffentlichungen der Europäischen Union.
- [3] M. Arndt, M. Baade, R. Berndt, F. Bunar, B. Bunel, G. Graf, M. Grubmüller, H. Jansen, D. Köhler, E.-M. Moser, L. Walter, R. Wanker, F. Schrade, H. Tschöke und K. Vidmar. *Real Drive Emissions (RDE): Gesetzgebung, Vorgehensweise, Messtechnik, Motorische Maßnahmen, Abgasnachbehandlung, Auswirkungen*. 1. Auflage 2019. ATZ/MTZ-Fachbuch. Springer Fachmedien Wiesbaden, 2018.
- [4] K. J. Åström und T. Hägglund. *Advanced PID control*. Research Triangle Park, NC: ISA - The Instrumentation Systems and Automation Soc, 2006.
- [5] B. Axelsson und P. O. Witze. „Qualitative laser-induced incandescence measurements of particulate emissions during transient operation of a TDI diesel engine“. In: *SAE Technical Paper Series*. SAE Technical Paper Series. SAE International 400 Commonwealth Drive, Warrendale, PA, United States, 2001.
- [6] *Excitation signals for nonlinear dynamic modeling of combustion engines*. Proceedings of the 17th world congress the international federation of automatic control Seoul, Korea, 2008. URL: <http://folk.ntnu.no/skoge/prost/proceedings/ifac2008/data/papers/1590.pdf>.
- [7] J. Belz. „Fighting the curse of dimensionality with local model networks: Bekämpfung des Fluchs der Dimensionalität mit lokalen Modellnetzen“. Dissertation. Universität Siegen, Institut für Mechanik und Regelungstechnik - Mechatronik, 2019.
- [8] B. Berger. „Modeling and optimization for stationary base engine calibration“. Dissertation. 2012.
- [9] S. A. Billings. *Nonlinear system identification: NARMAX methods in the time, frequency, and spatio-temporal domains*. Chichester, West Sussex: John Wiley & Sons, 2013.

-
- [10] C. M. Bishop. *Neural networks for pattern recognition*. Reprinted. Oxford: Oxford Univ. Press, 2010.
- [11] S. Björklund. *A survey and comparison of time-delay estimation methods in linear systems*. Bd. 1061. Linköping studies in science and technology. Thesis. Linköping: Univ, 2003.
- [12] C. Bohn und H. Unbehauen. *Identifikation dynamischer Systeme: Methoden zur experimentellen Modellbildung aus Messdaten*. Wiesbaden: Springer Vieweg, 2016.
- [13] J. D. Bomberger und D. E. Seborg. „Determination of model order for NARX models directly from input-output data“. In: *Journal of Process Control* 8.5-6 (1998), S. 459–468.
- [14] Bundesministerium für Wirtschaft und Energie. *Entwicklung der CO₂-Emissionen 1995 bis 2015 - weltweit und in Deutschland (in Millionen Tonnen): CO₂ Emissions from Fuel Combustion 2017 – Highlights*. Hrsg. von Internationale Energieagentur. 2017. URL: https://www.bmwk.de/Redaktion/DE/Downloads/I/Infografiken/infografik-klimaschutz-2.pdf?__blob=publicationFile&v=4.
- [15] K. Buß. „Testprozeduren für Lithium Ionen Batterien in (teil-) elektrifizierten Kraftfahrzeugen“. In: *Entscheidungen beim Übergang in die Elektromobilität*. Hrsg. von H. Proff. Wiesbaden: Springer Gabler, 2015, S. 317–333.
- [16] R. H. Byrd, J. C. Gilbert und J. Nocedal. *A trust region method based on interior point techniques for nonlinear programming*. Hrsg. von Springer-Verlag. 1996. URL: <https://hal.inria.fr/file/index/docid/73794/filename/RR-2896.pdf>.
- [17] R. H. Byrd, J. C. Gilbert und J. Nocedal. „A trust region method based on interior point techniques for nonlinear programming“. In: *Mathematical Programming* 89.1 (2000), S. 149–185.
- [18] R. H. Byrd, M. E. Hribar und J. Nocedal. „An interior point algorithm for large-scale nonlinear programming“. In: *SIAM Journal on Optimization* 9.4 (1999), S. 877–900.
- [19] M. Cabrera Cano. „Neuronale Netze mit externen Laguerre-Filtern zur automatischen numerischen Vereinfachung von Getriebemodellen“. Dissertation. URL: <http://nbn-resolving.de/urn:nbn:de:0072-644623>.
- [20] K. Chan, A. Saltelli und Tarantola Stefano. „Sensitivity analysis of model output: variance-based methods make the difference“. In: Bd. 29th Winter Simulation Conference Proceedings, S. 261–268.
- [21] M. Cieslik. „Analyse und Optimierung von Emissionen in RDE-Zyklen mittels P1-Hybridisierung eines Ottomotors: Analysis and optimization of RDE-cycle emissions for a P1-hybridization of a gasoline engine“. Master Thesis. Darmstadt: Technische Universität Darmstadt, 2017.

-
- [22] T. M. Cover und J. A. Thomas. *Elements of information theory*. 2nd ed. Hoboken, N.J: Wiley-Interscience, 2006.
- [23] E. Cramer und U. Kamps. *Grundlagen der Wahrscheinlichkeitsrechnung und Statistik: Ein Skript für Studierende der Informatik, der Ingenieur- und Wirtschaftswissenschaften*. 3., überarb. Aufl. Springer-Lehrbuch. Berlin: Springer Spektrum, 2014.
- [24] G. C. Critchfield, K. E. Willard und D. P. Connelly. „Probabilistic sensitivity analysis methods for general decision models“. In: *Computers and Biomedical Research* 19.3 (1986), S. 254–265.
- [25] R. I. Cukier, C. M. Fortuin, K. E. Shuler und Petschek, A.G. and Schaibly, J.H. „Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients: Part I: theory“. In: *Journal of chemical physics*, Vol. 59. Bd. Vol. 59, No. 8, S. 3873–3878.
- [26] C. Czado und T. Schmidt. *Mathematische Statistik*. Statistik und ihre Anwendungen. Berlin, Heidelberg: Springer-Verlag, 2011.
- [27] M. Deflorian. *Versuchsplanung und Methoden zur Identifikation zeitkontinuierlicher Zustandsraummodelle am Beispiel des Verbrennungsmotors: Zugl.: München, Techn. Univ., Diss., 2011*. Berichte aus der Automatisierungstechnik. Aachen: Shaker, 2012.
- [28] R. Dittmar. *Advanced process control: PID-Basisregelungen, Vermaschte Regelungsstrukturen, Softsensoren, Model predictive control*. München, Wien: De Gruyter Oldenbourg, 2017.
- [29] Y. Dursun. „Kennfeldoptimierung mit eingekoppelter Modellauswertung“. Diplomarbeit. KIT - Institut für Kolbenmaschinen, 2013.
- [30] Y. Dursun, S. Weber, R. Jakobi, F. Kirschbaum und S. Rinderknecht. „Approach for automated adjusting of the road load and tire simulation on powertrain test beds“. In: *Automotive data analytics, methods, DoE*. Hrsg. von K. Röpke und C. Gühmann. Bd. 9.4. Reihe Technik. Renningen: expert verlag, 2017.
- [31] N. Ebert, J.-C. Goos, F. Kirschbaum, E. Yildiz und T. Koch. „Methods of sensitivity analysis in model-based calibration“. In: *Automotive and Engine Technology* 5.1-2 (2020), S. 45–56.
- [32] N. Ebert, F. Kirschbaum und T. Koch. „Time-delay estimation for automotive applications“. In: *International Conference on Calibration Methods and Automotive Data Analytics*. Hrsg. von I. C. o. C. M. Analytics und A. Data. Tübingen: expert, 2019, S. 21–31.
- [33] A. Eggels und D. Crommelin. „Quantifying data dependencies with Rényi mutual information and minimum spanning Trees“. In: *Entropy* 21.2 (2019), S. 100.

-
- [34] J. Espinosa, J. Vandewalle und V. Wertz. *Fuzzy logic, identification and predictive control*. Advances in industrial control. London: Springer, 2005.
- [35] K.-T. Fang, R. Li und A. Sudjianto. *Design and modeling for computer experiments*. Bd. 6. Computer science and data analysis series. Boca Raton, Fla.: Chapman & Hall/CRC, 2006.
- [36] C. Fortgens. *On time delay estimation and its application to the study of communication between different cerebral areas*. Hrsg. von Computer Science. 1984.
- [37] J. H. Friedman. „Greedy function approximation: a gradient boosting machine“. In: *The Annals of Statistics* 29.5 (2001), S. 1189–1232.
- [38] M. H. Froehlich. „Informationstheoretische Optimierung künstlicher neuronaler Netze : für den Einsatz in Steuergeräten“. Dissertation. Tübingen: Universität Tübingen, 2003.
- [39] I. M. Gelfand und A. M. Jaglom. „Die Integration in Funktionenräumen und ihre Anwendung in der Quantentheorie“. In: *Fortschritte der Physik* 5.10-11 (1957), S. 517–556.
- [40] D. Gencaga, N. K. Malakar und D. J. Lary. „Survey on the estimation of mutual information methods as a measure of dependency versus correlation analysis“. In: AIP Conference Proceedings. AIP Publishing LLC, 2014, S. 80–87.
- [41] Geritt Kampmann, Nastasa Kieft und Oliver Nelles. „Support vector machines for design space exploration“. In: *Lecture notes in engineering and computer science 2. international association of engineers*. HongKong: Newswood Ltd, S. 1116–1121.
- [42] C. Gletter. *Künstliche Neuronale Netze Toolbox für Matlab, Daimler AG*. Version 0.20.2. 2020.
- [43] D. E. Goldberg. *Genetic algorithms in search, optimization, and machine learning*. 30. print. Boston: Addison-Wesley, 2012.
- [44] A. Goldstein, A. Kapelner, J. Bleich und E. Pitkin. *Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation*. Hrsg. von Cornell University. 2014.
- [45] M. Hafner. „Dynamische Motorvermessung“. In: *Modellgestützte Steuerung, Regelung und Diagnose von Verbrennungsmotoren*. Hrsg. von R. Isermann. Berlin und Heidelberg: Springer, 2003, S. 153–162.
- [46] M. Hafner. *Modellbasierte stationäre und dynamische Optimierung von Verbrennungsmotoren am Motorenprüfstand unter Verwendung neuronaler Netze: Zugl.: Darmstadt, Techn. Univ., Diss. Als Ms. gedr.* Bd. 482. Berichte aus dem Institut für Automatisierungstechnik der TU Darmstadt. Düsseldorf: VDI-Verl., 2002.

-
- [47] S. O. Han. „Varianzbasierte Sensitivitätsanalyse als Beitrag zur Bewertung der Zuverlässigkeit adaptiver Struktursysteme“. Dissertation. Darmstadt: Technische Universität Darmstadt, 2011.
- [48] B. Hartmann. „Lokale Modellnetze zur Identifikation und Versuchsplanung nichtlinearer Systeme“. Dissertation. Universität Siegen, 2014.
- [49] X. He und H. Asada. „A new method for identifying orders of input-output models for nonlinear dynamic systems“. In: *1993 American Control Conference*. IEEE, 2.06.1993 - 04.06.1993, S. 2520–2523.
- [50] A. Heinz, M. Moradi, H. Kubach und T. Koch. „Forschungsarbeit zum experimentellen Einsatz von NO_x-Sensoren für transiente Prüfläufe“. Projektarbeit. Karlsruhe: KIT, 2020.
- [51] A. O. Hero III und O. Michel. „Robust entropy estimation strategies based on edge weighted random graphs“. In: *Bayesian Inference for Inverse Problems*. Hrsg. von A. Mohammad-Djafari. SPIE Proceedings. SPIE, 1998, S. 250–261.
- [52] C. Hintz. „Identifikation nichtlinearer mechatronischer Systeme mit rekurrenten Netzen“. Dissertation. Technische Universität München und Lehrstuhl für Elektrische Antriebssysteme, 2003.
- [53] J. H. Holland. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor: Univ. of Michigan Press, 1975.
- [54] K. Hornik, M. Stinchcombe und H. White. „Multilayer feedforward networks are universal approximators“. In: *Neural Networks 2* (1989), S. 359–366.
- [55] IAV automotive engineering. *IAV Kasai and Online-DoE*.
- [56] S. Ishizuka, I. Kajiwara, J. Sato und Y. Hanamura. „A new approach for NO_x soft sensors for the aftertreatment of diesel engines“. In: *Journal of Physics: Conference Series 744* (2016), S. 012207.
- [57] J. Kennedy und R. Eberhart. „Particle swarm optimization“. In: *Proceedings of ICNN'95 - International Conference on Neural Networks*. IEEE, 27 Nov.-1 Dec. 1995, S. 1942–1948.
- [58] U. Kiencke. *Signale und Systeme*. München: Oldenbourg, 1998.
- [59] F. Kirschbaum. *Modellbasierte Applikationsverfahren: Empirische Methoden der Modellbildung zur modellbasierten Optimierung von Steuergeräteparametern in PKW-Antriebssträngen*. Stuttgart, 2017.
- [60] J. P. C. Kleijnen. *Design and analysis of simulation experiments*. 2nd ed. 2015. Bd. 230. International Series in Operations Research & Management Science. Cham: Springer, 2015.
- [61] W. Kleppmann. *Taschenbuch Versuchsplanung: Produkte und Prozesse optimieren*. 7., aktualisierte u. erw. Aufl. Praxisreihe Qualitätswissen. 2011.

-
- [62] K. Knödler. *Methoden der restringierten Online-Optimierung zur Basisapplikation moderner Verbrennungsmotoren: Zugl.: Tübingen, Univ., Diss., 2004*. Berlin: Logos, 2004.
- [63] K. Kroschel, G. Rigoll und B. Schuller. *Statistische Informationstechnik: Signal - und Mustererkennung, Parameter- und Signalschätzung*. 5. Aufl. Berlin, Heidelberg: Springer-Verlag, 2011.
- [64] M. Langner. *Impuls: Stickstoff und Verkehr: Journalistenworkshop "Stickstoff und Verkehr: mehr als ein Diesel-Problem": Fachgebiet II 4.1 / Grundsatzfragen der Luftreinhaltung*. Hrsg. von Umweltbundesamt. 2018. URL: https://www.umweltbundesamt.de/sites/default/files/medien/2546/dokumente/presentation_langner-veroeffentlichung.pdf.
- [65] J. Liebl, F. Munk, H. Hohenner und B. Ludwig. „Die Steuerung der neuen BMW Valvetronic-Motoren“. In: *MTZ - Motortechnische Zeitschrift* 62.7-8 (2001), S. 516–527.
- [66] I. Lind. *Regressor and structure selection: Uses of ANOVA in system identification*. Bd. 1012. Linköping studies in science and technology. Dissertations. Linköping: Linköpings universitet, 2006.
- [67] I. Lind. *Regressor selection in system identification using ANOVA*. Bd. No. 921. Linköping studies in science and technology. Linköping: Department of Electrical Engineering, Linköpings Universitet, 2001.
- [68] Linssen Rene und Henrich Christian. „Verbesserte DoE-Optimierungen und Verbrauchsvorhersagen: Zyklusreduktion mit einem erweiterten kmeans Verfahren“. In: *Design of Experiment in Engine Developments IAV 6* (2011), S. 183–204.
- [69] R. Linssen. „Modellbasierte Analyse und Optimierung von Dieselmotoren“. Dissertation. Aachen: Techn. Hochschule, 2010.
- [70] R. Linssen und C. Henrich. „Cluster analysis using an extended k-means process“. In: *Design of experiments (DoE) in engine development*. Hrsg. von K. Röpke. Renningen: Expert-Verl., 2011, S. 183–204.
- [71] R. Linssen, F. Uphaus und J. Mauss. „Simulation vernetzter Steuergeräte für die Fahrbarkeitsapplikation“. In: *ATZelektronik* 11.4 (2016), S. 16–21.
- [72] L. Ljung. *System identification: Theory for the user*. 2. ed., 14. printing. Prentice Hall information and system sciences series. Upper Saddle River, NJ: Prentice Hall PTR, 2012.
- [73] L. Ljung und T. Glad. *Modeling of dynamic systems*. Englewood Cliffs, N.J: PTR Prentice Hall, 1994.
- [74] N. Lüdtke, S. Panzeri, M. Brown, D. S. Broomhead, J. Knowles, M. A. Montemurro und D. B. Kell. „Information-theoretic sensitivity analysis: a general method for credit assignment in complex networks“. In: *Journal of the Royal Society, Interface* 5.19 (2008), S. 223–235.

-
- [75] D. J. C. MacKay. *Information theory, inference, and learning algorithms*. [Nachdr.] Cambridge: Cambridge Univ. Press, 2011.
- [76] R. Mannale. *Comparison of Regressor Selection Methods in System Identification: Technical Report*. Bd. LiTH-ISY-R-2730. 2006.
- [77] N. Mars und G. van Arragon. „Time delay estimation in nonlinear systems“. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 29.3 (1981), S. 619–621.
- [78] H. Maschmeyer. *Systematische Bewertung verbrennungsmotorischer Antriebssysteme hinsichtlich ihrer Realfahrtemissionen am Motorenprüfstand*. [1. Auflage]. Bd. Band 9. Schriftenreihe des Instituts für Verbrennungskraftmaschinen und Fahrzeugantriebe. Aachen: Shaker Verlag, 2018.
- [79] M. D. McKay, R. J. Beckman und W. J. Conover. „Comparison of three methods for selecting values of input variables in the analysis of output from a computer code“. In: *Technometrics* 21.2 (1979), S. 239–245.
- [80] J. Merkisz und J. Pielecha. „Comparison of Real Driving Emissions tests“. In: *IOP Conference Series: Materials Science and Engineering* 421 (2018), S. 042055.
- [81] H.-J. Mittag. *Statistik: Eine Einführung mit interaktiven Elementen*. 4., wesentlich überarbeitete und erweiterte Auflage. Springer-Lehrbuch. Berlin und Heidelberg: Springer Spektrum, 2016. URL: <http://dx.doi.org/10.1007/978-3-662-47132-6>.
- [82] D. C. Montgomery. *Design and analysis of experiments*. 8. ed. Hoboken, NJ: Wiley, 2013.
- [83] M. Mrosek, H. Sequenz und R. Isermann. „Control oriented NO_x and soot models for diesel engines“. In: *IFAC Proceedings Volumes* 43.7 (2010), S. 234–239.
- [84] T. Naumann. „Wissensbasierte Optimierungsstrategien für elektronische Steuergeräte an Common-Rail-Dieselmotoren“. Dissertation. Technische Universität Berlin, 2002.
- [85] O. Nelles. *Nonlinear system identification: From classical approaches to neural networks and Fuzzy models*. Berlin und Heidelberg: Springer, 2001. URL: <http://dx.doi.org/10.1007/978-3-662-04323-3>.
- [86] A. Neßler. *Optimierungsstrategien in der modellbasierten Dieselmotorenentwicklung*. Dissertation. Berlin, 2015.
- [87] D. Neumann, M. Steinbach, T. Kutzner, A. Lehmann, V. Kassem und M. Dreiser. „Model supported calibration process for future RDE requirements“. In: *16. Internationales Stuttgarter Symposium: Automobil- und Motorentechnik*. Hrsg. von M. Bargende, H.-C. Reuss und J. Wiedemann. Wiesbaden: Springer Fachmedien Wiesbaden, 2016, S. 565–579.

-
- [88] M. Nørgaard. *Neural networks for modelling and control of dynamic systems: A practitioner's handbook*. 3. printing, with corrections. Advanced textbooks in control and signal processing. London: Springer, 2003.
- [89] L. Pagel. *Information ist Energie: Definition eines physikalisch begründeten Informationsbegriffs*. Wiesbaden: Springer Vieweg, 2013.
- [90] M. Papageorgiou, M. Leibold und M. Buss. *Optimierung: Statische, dynamische, stochastische Verfahren für die Anwendung*. 4., korrigierte Auflage. Berlin und Heidelberg: Springer Vieweg, 2015.
- [91] J. Pillas. *Modellbasierte Optimierung dynamischer Fahrmanöver mittels Prüfständen*. 1. Auflage. Forschungsberichte Mechatronische Systeme im Maschinenbau. Aachen: Shaker, 2017.
- [92] J. Poland. *Modellgestützte und evolutionäre Optimierungsverfahren für die Motorentwicklung: Zugl.: Tübingen, Univ., Diss., 2002*. Berlin: Logos-Verl., 2002.
- [93] B. Pompe. „Measuring statistical dependences in a time series“. In: *J. STAT. PHYS* 73 (1993), S. 587–610.
- [94] B. Pompe. „Ranking and entropy estimation in nonlinear time series analysis“. In: *Nonlinear Analysis of Physiological Data*. Hrsg. von H. Kantz, J. Kurths und G. Mayer-Kress. Berlin, Heidelberg: Springer Berlin Heidelberg, 1998, S. 67–90.
- [95] W. H. Press. *Numerical recipes: The art of scientific computing*. 3rd ed. Cambridge, UK und New York: Cambridge University Press, 2007.
- [96] C. E. Rasmussen und C. K. I. Williams. *Gaussian processes for machine learning*. 3. print. Adaptive computation and machine learning. Cambridge, Mass.: MIT Press, 2008.
- [97] K. Reif, Hrsg. *Dieselmotor-Management im Überblick: Einschließlich Abgastechnik*. 2. Aufl. Bosch Fachinformation Automobil. Wiesbaden: Springer Vieweg, 2014.
- [98] J. Rettberg. „Implementierung einer robusten Methode zur automatisierten Totzeitkorrektur bei der Ermittlung von Emissionswerten: Implementation of a robust method for an automated time-delay adjustment in emissions value determination“. Masterarbeit. Stuttgart: Universität Stuttgart, September 2020.
- [99] J. Riegel. „Exhaust gas sensors for automotive emission control“. In: *Solid State Ionics* 152-153 (2002), S. 783–800.
- [100] D. E. Rivera, H. Lee, M. W. Braun und H. D. Mittelmann. „Plant-Friendly" system identification: a challenge for the process industries“. In: *IFAC Proceedings Volumes* 36.16 (2003), S. 891–896.

-
- [101] F. Rodríguez, Y. Bernard, J. Dornoff und P. Mock. *Recommendations for post-euro 6 standards for light-duty vehicles in the European Union*. Hrsg. von International Council on Clean Transportation. 2019. URL: <https://theicct.org/publications/recommendations-post-euro-6-eu>.
- [102] A. Saltelli. *Sensitivity analysis in practice: A guide to assessing scientific models*. Reprinted. Hoboken, NJ: Wiley, 2007.
- [103] R. Sauermann, D. Boja, K. Frank und O. Nelles. „Particle swarm optimization for automotive model-based calibration“. In: *6th IFAC Symposium Advances in Automotive Control*, S. 733–738.
- [104] E. Schaffernicht, R. Kaltenhaeuser, S. S. Verma und H.-M. Gross. „On estimating mutual information for feature selection“. In: *Artificial Neural Networks – ICANN 2010*. Bd. 6352. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, S. 362–367.
- [105] S. Scheidel, M.-S. Grande, G. Zerbinì und M. Decker. „A versatile approach for transient maneuver optimization using DoE methods“. In: *International Conference on Calibration Methods and Automotive Data Analytics*. Hrsg. von I. C. o. C. M. Analytics und A. Data. Tübingen: expert, 2019, S. 219–231.
- [106] R. Schlittgen. *Multivariate Statistik*. Oldenbourg Wissenschaftsverlag, 2011, S. I–XII.
- [107] M. Schmidt. „An interior-point method for nonlinear optimization problems with locatable and separable nonsmoothness“. In: *EURO Journal on Computational Optimization* 3.4 (2015), S. 309–348.
- [108] M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo und H.-P. Schwefel, Hrsg. *Parallel Problem Solving from Nature PPSN VI: 6th International Conference Paris, France, September 18-20, 2000 Proceedings*. Bd. 1917. Lecture Notes in Computer Science. Berlin und Heidelberg: Springer, 2000.
- [109] C. A. Scholbeck, C. Molnar, C. Heumann, B. Bischl und G. Casalicchio. „Sampling, intervention, prediction, aggregation: A generalized framework for model-agnostic interpretations“. In: *Machine Learning and Knowledge Discovery in Databases*. Hrsg. von P. Cellier und K. Driessens. Bd. 1167. Communications in Computer and Information Science. Springer International Publishing, 2020, S. 205–216.
- [110] D. Scholz. *Optimierung interaktiv: Grundlagen verstehen, Modelle erforschen und Verfahren anwenden*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018.
- [111] D. Schröder. *Intelligente Verfahren: Identifikation und Regelung nichtlinearer Systeme*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg, 2010.

-
- [112] M. Schröder, N. Baltes und J. Danzer. „Potenzial von Prüfstandsmessungen unter RDE-Anforderungen“. In: *MTZ – Motortechnische Zeitschrift* 78.6 (2017), S. 32–41.
- [113] V. Schwieger. *Nicht-lineare Sensitivitätsanalyse gezeigt an Beispielen zu bewegten Objekten: Zugl.: Stuttgart, Univ., Habil.-Schr., 2004*. Bd. 581. Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften Reihe C, Dissertationen. München: Beck, 2005.
- [114] C. E. Shannon. „A mathematical theory of communication“. In: *The Bell System Technical Journal*. Bd. 27. 1948, S. 379–423. URL: <http://www.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>.
- [115] K. Siebertz, D. van Bebbber und T. Hochkirchen. *Statistische Versuchspaltung: Design of Experiments (DoE)*. 2nd ed. VDI-Buch Ser. Berlin, Heidelberg: Vieweg, 2017.
- [116] A. Sittig. „Optimierung und Applikation von Betriebsstrategien in Hybridfahrzeugen“. Dissertation. Technische Universität München, 2014.
- [117] I. M. Sobol. „Sensitivity estimates for nonlinear mathematical models“. In: *Matem. Modelirovanie* 2 (1) 1990, S. 112–118.
- [118] . Staaten und E. Union. *Paris Agreement: Ratification and accession by 55 UNFCCC parties, accounting for 55% of global greenhouse gas emissions*. Paris, 4 November 2016. URL: https://unfccc.int/sites/default/files/english_paris_agreement.pdf.
- [119] J. Suzuki. „An estimator of mutual information and its application to independence testing“. In: *Entropy* 18.4 (2016), S. 109.
- [120] A. K. Tangirala. *Principles of system identification: Theory and practice*. Boca Raton, FL: CRC Press, 2015.
- [121] M. Temmler. „Steuergerätetaugliche Verbrennungsoptimierung mit physikalischen Modellansätzen“. Dissertation.
- [122] I. The MathWorks. *Optimization Toolbox - User's Guide*. Hrsg. von I. The MathWorks. 2020. URL: https://www.mathworks.com/help/pdf_doc/optim/optim.pdf.
- [123] Umweltbundesamt. *Treibhausgas-Emissionen seit 1990 nach Gasen: Nationale Treibhausgas-Inventare 1990 bis 2019 (Stand 12/2020) sowie Vorjahres-schätzung (VJS) für das Jahr 2020 (PI 07/2021 vom 15.03.2021)*. Hrsg. von Umweltbundesamt. 2021. URL: https://www.umweltbundesamt.de/sites/default/files/medien/384/bilder/dateien/2_abb_thg-emissionen-seit-1990-nach-gasen_2021_0.pdf.
- [124] R. J. Vanderbei und D. F. Shanno. „An interior-point algorithm for nonconvex nonlinear programming“. In: *Computational Optimization and Applications* 13.1/3 (1999), S. 231–252.

-
- [125] R. A. Waltz, J. L. Morales, J. Nocedal und D. Orban. „An interior algorithm for nonlinear optimization that combines line search and trust region steps“. In: *Mathematical Programming* 107.3 (2006), S. 391–408.
- [126] M. Werner. *Information und Codierung: Grundlagen und Anwendungen*. 2. Aufl. Wiesbaden: Vieweg + Teubner, 2009.
- [127] A. Westlund. „Measurement of diesel engine emissions during transient operation“. Master Thesis. Stockholm: KTH Industrial Engineering and Management, 2007. URL: <https://www.diva-portal.org/smash/get/diva2:544701/fulltext01.pdf>.
- [128] H. Winner, T. Koch, F. Gauterin und H. Rottengruber. *Aktuelle Entwicklungen zur „CO₂- Gesetzgebung im Verkehrssektor“*. Hrsg. von Wissenschaftlichen Gesellschaft für Kraftfahrzeug- und Motorentechnik e.V. 2018. URL: https://www.wkm-ev.de/images/2018-12-20_Positionspapier-aktueller-CO2-Gesetzentwurf-Fahrzeugantriebe.pdf.
- [129] H. Winner, J. Wiedemann, F. Gauterin und H. Rottengruber. *Die Zukunft des Verbrennungsmotors: Bewertung der dieselmotorischen Situation*. Hrsg. von Wissenschaftlichen Gesellschaft für Kraftfahrzeug- und Motorentechnik e.V. 2017. URL: <https://www.wkm-ev.de/de/aktuelles/21-statement-die-zukunft-des-verbrennungsmotors-bewertung-der-dieselmotorischen-situation.html>.
- [130] S. Zydek. „Optimale Regelung des Dieselmotors zur Minimierung von Realfahrtemissionen und -verbrauch“. Dissertation. Darmstadt: Technische Universität Darmstadt, 2018. URL: <https://tuprints.ulb.tu-darmstadt.de/8068/>.
- [131] R. van Basshuysen und F. Schäfer, Hrsg. *Handbuch Verbrennungsmotor*. Wiesbaden: Springer Fachmedien Wiesbaden, 2015.
- [132] P. van Overschee und B. de Moor. „N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems“. In: *Automatica* 30.1 (1994), S. 75–93.

Nomenklatur

In Anlehnung an die national und international gebräuchliche Nomenklatur sind die jeweiligen Variablen, Operatoren, Konstanten und Kennzahlen aufgeführt. Die Abkürzungen orientieren sich dabei vorwiegend an der englischen Sprache der Fachliteratur und sind *kursiv* gekennzeichnet. Das Dezimaltrennzeichen ist das Komma. Die Ergebnisse sind vorwiegend normiert dargestellt.

Lateinische Zeichen

Zeichen	Einheit	Bedeutung
a_i	-	Abstandswert
$A(q), B(q), \dots$	-	Polynome der Systemidentifikation
$\mathbf{A}, \mathbf{B}, \mathbf{C}$	-	Gleichverteilte Zufallsmatrizen, Systemmatrix
\mathbf{b}	-	Eingangsvektor
b_{eff}	g kWh ⁻¹	Effektiver spezifischer Kraftstoffverbrauch
c	-	Design Parameter
\mathbf{c}^\top	-	Ausgangsvektor
C	-	Schätzwert der Entropie
c_i	%	Stoffkonzentration
d	-	Dimension, Grad, Durchgriff
E	-	Fehlerfunktion
$e(\cdot)$	-	Externes Störsignal
f	Hz	Frequenz
f_s	Hz	Abtastfrequenz
$f(\cdot)$	-	Funktion
g_k	-	Globaler Bestwert eines Partikels
$g(\cdot)$	-	Gleichungsnebenbedingung
h	-	Erfahrungsgewicht eines Partikels
$h(\cdot)$	-	Ungleichungsnebenbedingung
I	bit	Transinformation
k	-	Diskreter Zeitschritt
k_{SE}	-	Quadratisch exponentieller Kernel
\mathbf{K}	-	Kovarianzmatrix
l	-	Stufen/Level der Versuchsplanung

Zeichen	Einheit	Bedeutung
$l_{i,j}$	-	Eindimensionaler Lipschitz-Quotient
L	-	Lipschitz-Quotient
ℓ	-	Längenskala
m	kg	Masse
$m_{i,j}$	-	Gewichtung
\dot{m}	g s^{-1}	Massenstrom
M	Nm	Drehmoment, Last
\mathbf{M}	-	Interpolationsmatrix
\dot{M}	Nm s^{-1}	Lastgradient
n	min^{-1}	(Motor-) Drehzahl
n_a, n_b, n_c, \dots	-	Ordnung Zähler- oder Nennerpolynom
n_k	-	Verschiebeoperator/ Verzögerungsglied
n_M	-	Anzahl an Basisfunktionen
n_m	-	Modellordnung eines dynamischen Modells
n_p	-	Anzahl der Dimension
n_q	-	Anzahl der Quotienten
N, M, L, K	-	Anzahl Messdaten oder Signallänge
o_m	-	Anzahl der Pole
o	-	Anzahl identischer Werte
p_{ALD}	bar	Abgasladedruck
p_{eff}	bar	Effektiver Mitteldruck
p_{Rail}	bar	Raildruck
p_k^i	-	Bestwert eines Partikel
$p(\cdot)$	-	Wahrscheinlichkeitsverteilung
q_{MI}	mg Hub^{-1}	Einspritzmenge der Haupteinspritzung
$q_{\text{I, Ges}}$	mg Hub^{-1}	Gesamteinspritzmenge Einspritzung
q_{Post}	mg Hub^{-1}	Einspritzmenge der Nacheinspritzung
r	-	Relativer Rang, Faktor, Zufallszahl
r_p	-	Korrelationskoeffizient nach Pearson
r_{yu}	-	Kreuzkorrelationsfunktion
$r : , s : , t :$	-	Rang- Abbildungsfolgen
$R(\cdot)$	-	Rang/ Rangfolge
\mathbf{R}	-	Matrix der Glattheitsmaße
s	km	Strecke
s_u	-	Schätzwert der Standardabweichung
s_u^2	-	Schätzwert der Varianz
s_{uy}	-	Schätzwert der Kovarianz
S	-	Sensitivitätsindex
t	s	Zeit
T_{10-90}	s	Ansprech-/ Anstiegszeit

Zeichen	Einheit	Bedeutung
T_s	s	Abtastzeit
T_d	s	Totzeit
$\mathbf{u} = \{u_k\}_{k=1}^N$	-	Eingangswerte eines Systems
$u(\cdot)$	-	Zeit-/ Messreihe
$\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_{n_p}]^\top$	-	Menge von n_p -Zeitreihen
v_k	-	Geschwindigkeit eines Partikel
$v_i = [t, u]$	-	Kennfeldkoordinaten eines Betriebspunktes
w	-	Gewichtung oder Bias
x	-	Optimierungsgröße
x_i	-	Aktivität eines Neurons
$\mathbf{x}, \dot{\mathbf{x}}$	-	Zustandsvektoren
$\mathbf{y} = \{y_k\}_{k=1}^M$	-	Zeitdiskrete (Soll-)Werte/ Trainingsmenge
$y(\cdot)$	-	Zeit- oder Messreihe
$\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_{n_p}]^\top$	-	Menge von n_p -Zeitreihen
z	-	Schiebeoperator der Z-Transformation
$z_i = [x, y]$	-	Kennfeldkoordinaten einer Stützstelle

Mathematische Operatoren

Zeichen	Bedeutung
$\dot{(\cdot)}$	Gradient
$\hat{(\cdot)}$	Geschätzte Größe
\mathbf{u}	Vektor u
\mathbf{U}	Matrix U
$\overline{(\cdot)}$	Mittelwert eines Vektors
$(\cdot)^+$	Nichtdimensionale Schrittweite
$(\cdot)^\top$	Transponierte einer Matrix
$(\cdot)^{-1}$	Inverse einer Matrix
$ (\cdot) $	Determinante einer Matrix
∇	Nabla-Operator
\mathcal{B}	Binäre Rang-Abstands-Matrix
$\text{Cov}\{\cdot\}$	Kovarianzmatrix
$\mathbb{E}\{\cdot\}$	Erwartungswert
$V\{\cdot\}$	Varianz
\mathbf{J}	Jakobi-Matrix
$\mathcal{N}(\cdot)$	Gaußverteilung
$\mathcal{L}(\cdot)$	Lagrange-Funktion
Σ	Summenzeichen

Griechische Zeichen

Zeichen	Einheit	Bedeutung
α	-	Parameter der Reny-Entropie
α_z, α_λ	-	Schrittweite der Optimierung
β	-	Koeffizienten eines Modells
$\gamma_{95\%}$	-	95 %-Konfidenzintervall
δ_{d,kk^*}	-	Abstand der Punktepaare
Δ	-	Differenz
ε	-	Vergrößerungsniveau
ζ	%	Schiebeoperator der Kennfeldglattheit
η	-	Pedalwert
H	bit	Informationstheoretische Entropie
θ	-	Parameter
Θ	-	Gewichtung
λ	-	Lagrange-Multiplikator
μ	-	Mittelwertsvektor, Barriereparameter
ξ	-	Trägheitskonstante
ρ_i	g dm^{-3}	Dichte
ρ_{SP}	-	Korrelationskoeffizient nach Spearman
σ^2	-	Varianz
Σ_u	-	Kovarianzmatrix der Datenmenge \mathbf{U}
ϕ_{MI}	°KW	Einspritzzeitpunkt der Haupteinspritzung
φ	-	Regressor
Φ	-	Basis-/ Gültigkeitsfunktion
ψ_{AGR}	%	Rate der Abgasrückführung
ψ_{VTG}	%	Stellgröße der variablen Turbinengeometrie

Chemische Komponenten

Zeichen	Bedeutung
CO	Kohlenstoffmonoxid
CO ₂	Kohlenstoffdioxid
HC	Kohlenwasserstoff
N ₂ O	Distickstoffoxid
NO	Stickstoffoxid
NO ₂	Stickstoffdioxid
NO _x	Sammelbegriff für Stickstoffoxide

Hoch- und tiefgestellte Indizes

Zeichen	Bedeutung
0	Referenz
BP	Betriebspunkt
d	Derivat
DE	Direkteinspritzer
ex	Exakt
exh	Abgasmassenstrom
FAST	<i>Fourier amplitude sensitivity test</i>
FIT	Fitness/ Fehlerwert der Messdaten
gr	Größer
ident	Identisch
kl	Kleiner
KP	Kennfeld Stützstelle
L	Linear
LS	Methode kleinster Quadrate (<i>Least square</i>)
m	Modell
max	Maximalwert
min	Minimalwert
n	Rauschen (<i>Noise</i>)
nrm	Normierung
NL	Nichtlinear
OM	Ölmotor
Opt	Optimierung
PN	Partikelanzahl
Rau	Rauigkeit
SE	Quadratisch-exponentielle Kernel-Funktion
Stat	Stationäre Basisapplikation
TEMP	Temporär
VAL	Validierung
ZF	Zielfunktion oder -größe

Abkürzungen

Zeichen	Bedeutung
4SID	<i>State-space subspace system identification</i>
AMA	Abgasmessanlage
AIC	<i>Akaike information criterion</i>
APRBS	<i>Amplitude modulated pseudorandom bit sequences</i>
ARMAX	<i>Autoregressive moving average with exogenous input</i>
ARX	<i>Autoregressive with exogenous input</i>
CD	<i>Crowding distance</i>
CF	Konformitätsfaktor
CLD	Chemilumineszenz Detektor
DoE	Statistische Versuchsplanung (<i>Design of experiments</i>)
EOP	Position nach der Abgasanlage (<i>End of pipe</i>)
EU	Europäische Union
EZMI	Einspritzzeitpunkt der Haupteinspritzung
FANOVA	<i>Functional analysis of variance</i>
FAST	<i>Fourier amplitude sensitivity test</i>
FIR	<i>Finite impulse response</i>
FSN	Schwärzungszahl (<i>Filter smoke number</i>)
FTIR	<i>Fourier transformations infrarotspektrometer</i>
GMO	Genetischer multikriterieller Optimierungsansatz
GP(M)	Gaußprozess (Modell)
HDMR	<i>High dimensional model representation</i>
HILOMOT	<i>Hierarchical local model tree</i>
ICE	<i>Individual conditional expectation</i>
IIR	<i>Infinite impulse response</i>
IV	Instrumentenvariablen
KKT	Karush-Kuhn-Tucker-Bedingung
KNN	Künstliches Neuronales Netz
KW	Kurbelwinkel
LMA	Levenberg-Marquardt-Algorithmus
LMN	Lokales Modellnetz
LOLIMOT	<i>Local linear model tree</i>

Zeichen	Bedeutung
LS	Methode kleinster Quadrate (<i>Least square</i>)
ML	<i>Marginal likelihood</i> -Methode
MLP	<i>Multilayer perceptron</i>
MSE	Mittlerer quadratischer Fehler (<i>Mean square error</i>)
NDUV	<i>Non-dispersive ultra violet spectroscopy</i>
NARMAX	<i>Nonlinear ARMAX</i>
NARX	<i>Nonlinear ARX</i>
NEDC	<i>New European driving cycle</i>
NFIR	<i>Nonlinear FIR</i>
NOE	<i>Nonlinear OE</i>
NRMSE	Normalisierte Wurzel des MSE
NSGA-II	<i>Non-dominated sorting genetic algorithm</i>
OE	<i>Output error</i>
PD(P)	<i>Partial dependence</i> (plot)
PEMS	<i>Portable emission measurement system</i>
PKW	Personenkraftwagen
PN	Partikelanzahl (<i>Particle number</i>)
PRESS	Quadratsumme der Vorhersagefehler
PSO	Partikelschwarmoptimierung
RBF	<i>Radial basis function</i>
RDE	<i>Real driving emissions</i>
RMSE	Wurzel des MSE
RRL	<i>Real time-recurrent-learning</i>
TDE	<i>Time delay estimation</i>
TDNN	<i>Time delay neural network</i>
u. B. v.	Unter Berücksichtigung von
USA	Vereinigte Staaten von Amerika
WKM	Wissenschaftliche Gesellschaft für Kraftfahrzeug- und Motorentchnik e.V.
WLTC	<i>Worldwide harmonized light duty vehicle test cycle</i>

Abbildungsverzeichnis

1.1	Die Euro Grenzwertentwicklung und die Diskrepanz im realen Fahrverhalten	2
1.2	Struktur der Software eines Motorsteuergeräts	4
2.1	Allgemeiner Prozess der modellbasierten Kalibrierung	7
2.2	Lokales Modellnetz nach [85]	11
2.3	Unterscheidung von Simulator- und Prädiktor-Konfiguration im Modelltraining	14
2.4	Darstellung eines TDNN	15
2.5	Aufbau eines KNN aus überlagerten Basisfunktionen	16
2.6	Darstellung eines Perzeptron	18
2.7	Bilineare Interpolation eines Betriebspunktes	21
2.8	Schematische Darstellung einer Kennlinie des Problems mehrdeutiger Lösungen	22
3.1	Der Informationskanal nach [114]	29
3.2	Vergleich des Drehmomenten mit dessen relativer Rangfolge	32
3.3	Frequenzanteile der FFT-Analyse von Originalsignal und transformierter relativer Rangfolge	33
3.4	Zufallsverteilung von Datensätzen zum statistischen Vergleich von Korrelationsanalyse und Transinformation	34
3.5	Vergleich der Korrelationsanalyse der Kalibrierparameter auf die Ausgangsgrößen	35
3.6	Einfluss der Abtastzeit auf die Informationsmaße zweier verschiedener RDE-Ersatzzyklen	38
3.7	Überblick der Eingangsfunktionen des Anwendungsfalles (a) und der Einfluss der Stichprobengröße N auf den Schätzwert der Varianz des Gesamtmodells s_Y^2 (b)	44
3.8	Der GPM-Hyperparameter – der Längenskalar – variiert stark abhängig von der Datensatzgröße (a), während die Indizes der varianzbasierten Sensitivitätsanalyse S_{T_i} unabhängig von der Anzahl an Messungen und folglich der Modellgüte sind (b)	45
3.9	Ergebnisse der Sensitivitätsanalyse des generischen Anwendungsfalles mit $M = 200$ Punkten (a) und die zugehörigen Wahrscheinlichkeitsverteilungen $p(y_i)$ der verschiedenen Eingangsfunktionen y_i für eine Anzahl von 100 Punkten (b)	46

3.10	Ergebnisse der Sensitivitätsanalyse auf die Ausgangsgrößen NO_x - Massenstrom \dot{m}_{NO_x} (a) und effektiver Mitteldruck p_{eff} (b)	47
3.11	Visuelle Erklärbarkeitsansätze für Beispielmotive der Grundmoto- renkalibrierung	48
3.12	Lineare dynamische Modellstrukturen für die explizite Parameter- schätzung der Totzeiten	50
3.13	Vergleich der Signale zur TDE: Zufallssignal u , linear transformiertes Signal y_L und nichtlinear transformiertes Signal y_{NL}	54
3.14	Kreuzkorrelationen der linearen und nichtlinearen Signale als Funk- tion des Schiebeoperators κ	55
3.15	Vergleich des Modellfehlers MSE für ARX-Modellschätzungen bis zur Ordnung $n_m = 8$ für linearen (a) und nichtlinearen Anwendungsfall (b)	56
3.16	Ein Ausschnitt der normierten Signale des Abgasmassenstromes \dot{m}_{exh} und der Konzentration der Partikelanzahl c_{pN} für eine Berechnung des Abgaspartikel-Massenstroms in der Zertifizierung	57
3.17	Einfluss der Fahrzeugmasse pro Strecke s im WLTC auf die NO_x -Roh- (a) und EOP-Emissionen (b) normiert	59
3.18	Ergebnisse einer Clusterung des WLTC im Drehzahl-Last-Kennfeld mit 14 Centroiden und Leistungshyperbeln (a) und der Verweildauer des Zyklus in den jeweiligen Centroiden (b)	60
3.19	Abdeckung des Drehzahl-Last-Kennfeldes durch den synthetischen NEDC, den WLTC und einen möglichen RDE-Ersatzzyklus	61
3.20	Häufigkeitsverteilung für den Lastgradienten $\dot{M} = dM/dt$ eines WLTC (a) und eines RDE-Ersatzzyklus (b)	63
3.21	Exemplarischer Vorschlag einer neuen Definition des Betriebspunkts von einem zweidimensionalen Drehzahl-Last-Kennfeld (a) zu einem Wertetripel mit der Dynamik der ersten Ordnung, dem Lastgradienten, als dynamische Führungsgröße (b)	64
3.22	Interpolation eines Kennfeldes der normierten Rohemissionen von NO_x aus dem quasistationären Anteil $\dot{M} = [-4, 4]$ einer n - M - Bahnkurve eines RDE-Ersatzzyklus	66
3.23	Normierte Rohemissionen von NO_x entlang der n - M - \dot{M} -Bahnkurve eines RDE-Zyklus und eine interpolierte Ebene mit dem quasistatio- nären Anteil $\dot{M} = [-4, 4]$	66
3.24	Prinzipische Skizze der Interpolation einer transienten $f(n, M, \dot{M})$ - Bahnkurve auf eine stationäre Projektion im Kennfeld $f(n, M)$	67
3.25	Sprungantwort eines PT1-Gliedes mit Totzeit.	69
3.26	Vergleich der Rohemissionen des NO_x -Massenstroms gemessen mit dem NO_x -Sensor, AMA und FTIR auf die Sprünge der Gesamtein- spritzmenge $q_{1,\text{Ges}}$	70
4.1	Möglicher Versuchsumfang der Kalibrierparameter und Zielgrößen für eine Kalibrierungsaufgabe der Grundmotorenbedatung	72

4.2	Visuelle Erklärbarkeit der Eingangsgröße Einspritzzeitpunkt der Haupteinspritzung ϕ_{MI} auf die verschiedenen Ausgangsgrößen.	75
4.3	Analyse positiver Laständerungen der Anteile eines RDE-Ersatzzyklus im Drehzahl-Last-Kennfeld $n - M$ mit der Beschleunigungsdauer t (a) und deren zugehöriger Lastgradient mit kumulierter NO_x -Emission des Anteils über der gemittelten Drehzahl \bar{n} des Lastsprungs (b)	76
4.4	Vergleich der Istwertkurven des Drehmoments M der Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ bis $\dot{M} = 350 \text{ Nm s}^{-1}$ bei einer konstanten Drehzahl $n = 1600 \text{ min}^{-1}$	77
4.5	Vergleich der interpolierten Sollwertvorgabe des EZMIs ϕ_{MI} aus dem Kennfeld in Abhängigkeit der Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$ bis $\dot{M} = 350 \text{ Nm s}^{-1}$	78
4.6	Ausgewählte, geglättete Kennlinien der Versuchsplanung für die Optimierung des EZMIs ϕ_{MI} , aufgetragen über der Einspritzmenge q_{MI} bei konstanter Drehzahl n	79
4.7	Einfluss des Lastgradienten auf die Emissionen.	81
4.8	Einfluss von zehn Variationen des EZMIs ϕ_{MI} auf die Emissionsmassenströme \dot{m}_i von NO_x , CO und HC für den Lastgradienten $\dot{M} = 70 \text{ Nm s}^{-1}$	82
4.9	Streudiagramme der Modellprädiktionen über den Messwerten.	83
4.10	Ergebnisse der nichtlinearen dynamischen Modellierung.	85
4.11	Die Ergebnisse der Kennfeldstützstellen im Variationsbereich für die optimierten Kriterien im transienten Fahrbetrieb in Abhängigkeit von der Dynamik	87
4.12	Vergleich von vier Mittelwertmessungen der stationär optimierten Ursprungsparametrierung (schwarz) mit der optimierten Parametrierung (rot) für die Emissionsmassenströme von NO_x , CO und HC in Abhängigkeit vom Lastgradienten \dot{M}	88
4.13	Validierung der dynamischen Optimierung von Emissionen	89
4.14	Synthese der Erweiterung einer Motorsteuererätefunktion zu einer Kalibrierung des Wertetripels	91
A.1	Kurvenschar der Einspritzmenge q_{MI} sowie des Einspritzzeitpunktes der Haupteinspritzung ϕ_{MI} der Laständerungen abhängig vom Lastgradient \dot{M}	xvi
A.2	Kurvenschar der Emission-Massenströme \dot{m}_{CO_2} über die Laständerung abhängig vom Lastgradient \dot{M}	xvii
A.3	Kurvenschar der Emission-Massenströme \dot{m}_{NO_x} , \dot{m}_{CO} und \dot{m}_{HC} über die Laständerung abhängig vom Lastgradient \dot{M}	xvii
A.4	Vergleich der zweidimensionalen Pareto-Fronten der statischen Optimierung mittels GMO (a) und die Zielfunktionsauswertungen der PSO für die Optimierung der nichtlinearen dynamischen Modellierung über 300 Iterationen	xix

A.5	Streudiagramme der Modellprädiktionen \hat{y} über den Messwerten y der skalaren HC-Größen in Abhängigkeit vom Lastgradient \dot{M} für Validierungsdaten (Δ), Wiederholpunkte (+) und Messwerte (\bullet) . . .	xix
A.6	Darstellung der integralen NO_x -Massenstromkriterien der statischen Modellierung inklusive der Messausreißer (\circ) über der Variation sowie der Größe nach geordnet (+)	xx
A.7	Darstellung der integralen HC-Massenstromkriterien der statischen Modellierung inklusive der Messausreißer (\circ) über der Variation sowie der Größe nach geordnet (+)	xxi
A.8	Darstellung der integralen CO-Massenstromkriterien der statischen Modellierung inklusive der Messausreißer (\circ) über der Variation sowie der Größe nach geordnet (+)	xxii
A.9	ICE und PDP der NO_x , CO und HC-Modelle für den Lastgradient $\dot{M} = 70 \text{ Nm s}^{-1}$	xxiii
A.10	ICE und PDP der NO_x , CO und HC-Modelle für Lastgradient $\dot{M} = 117 \text{ Nm s}^{-1}$	xxiv
A.11	ICE und PDP der NO_x und CO-Modelle für $\dot{M} = 350 \text{ Nm s}^{-1}$	xxv
B.1	Iterative Bestimmung von ε_d zur Lösung der Zielfunktion $H_2(\mathbf{U}, \mathbf{Y}, \varepsilon_d) = -\text{ld}(0.05)$ mittels nichtlinearer Innere-Punkte-Optimierung	xxx
C.1	Exemplarischer Vergleich des vollfaktoriellen und modellbasierten Versuchsaufwands.	xxxvi
C.2	Beispiel der Iterationen eines HILOMOT-Ansatzes.	xxxix
C.3	Prozentuale Änderung des Lipschitz-Index	xli
C.4	Die Charakterisierung von Pareto-Fronten.	xliv
C.5	Zweidimensionale Visualisierung der Einteilung von Individuen in den jeweiligen Rang eines zweidimensionalen Zielfunktionsraums und schematische Bestimmung der CD für ein Individuum I_i	xlvii
C.6	Allgemeine Darstellung einer vollständigen linear dynamischen Modellstruktur für einen gestörten linearen Prozess nach [85, S. 467] . .	liii

Tabellenverzeichnis

1.1	Zeitleiste der Euronorm 6 und der zugehörigen Fahrzyklen.	2
3.1	Eigenschaften der verallgemeinerten Transinformation \hat{I}_2 für \mathbf{U} und \mathbf{Y} bei einer Gleichverteilung von \mathbf{Y} [nach 38, S. 31, 94, S. 75].	31
3.2	Statistische Maße zur Beurteilung der Datenmengen.	34
3.3	Kalibrierparameter einer modellbasierten Kalibrierungsaufgabe in der Grundmotorenkalibrierung	35
3.4	Eigenschaften der Sensitivitätsindizes [113, S. 33]	41
3.5	Zusammenfassung der Ordnungen und Ergebnisse der Schätzung der Totzeitglieder n_k der generischen Signale	55
3.6	Ergebnisse der Potenzialabschätzung für die Erweiterung eines Betriebspunkts zu einem Wertetripel	67
3.7	Verfügbare Messtechniken zur Betrachtung von NO_x -Emissionen.	69
4.1	Sensitivitätsanalyse für die Haupteinflüsse S_i der modellbasierten Kalibrierung mit den Kalibrierparametern u_i auf fünf Modellausgänge \hat{y}_i	74
4.2	Modellfehler der statischen Modellierung.	84
4.3	Validierte Parametrierung der Optimierungen der statischen und dynamischen Modellen des jeweiligen Lastgradienten sowie der Typ des numerischen Optimierungsalgorithmus	86
A.1	Ergebnisse der Optimierung der statischen Modelle und $\gamma_{95\%}$ Vertrauensbereich der Modelle	xviii
B.1	Vergleich der Rangtransformationen	xxix
C.1	Lösungsszenario für den Kennfeldvektor \hat{z} [30, S. 11]	lii

A Ergänzende Abbildungen der Optimierung des transienten Betriebs

Im folgenden Abschnitt sind die Abb. zur Prüfstandvermessung des Kap. 4 sowie die Optimierungsergebnisse der jeweiligen Variationen aufgeführt. Kurvenscharen der Einspritzmenge q_{MI} sowie des EZMIs ϕ_{MI} der Laständerungen abhängig vom Lastgradient \dot{M} sind in Abb. A.1 gezeigt. Die Emission-Massenströme von CO_2 sind in Abb. A.2 dargestellt. Erkennbar ist, dass die Linien sich ähneln und die Messtechnik kaum Variationen aufnehmen.

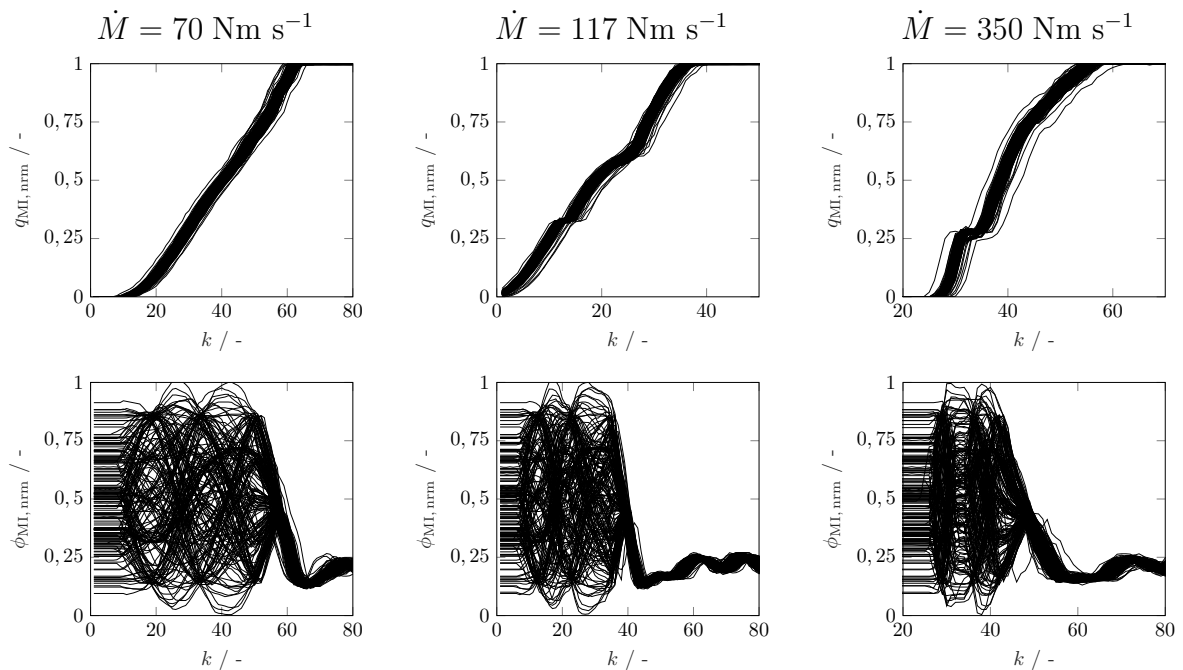


Abbildung A.1 Kurvenschar der Einspritzmenge q_{MI} sowie des Einspritzzeitpunktes der Haupteinspritzung ϕ_{MI} der Laständerungen abhängig vom Lastgradient \dot{M}

Kurvenscharen der Emission-Massenströme \dot{m}_{NO_x} , \dot{m}_{CO} und \dot{m}_{HC} über die Laständerung abhängig vom Lastgradient \dot{M} ist in Abb. A.3 gezeigt. Je nach Variation der Kennlinie der Haupteinspritzung kann dies die Steigung sowie Höhe des Signals beeinflussen.

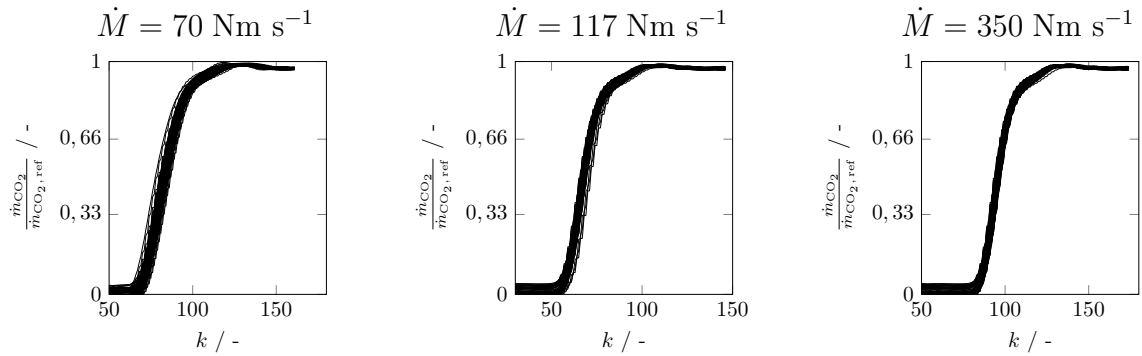


Abbildung A.2 Kurvenschar der Emission-Massenströme \dot{m}_{CO_2} über die Laständerung abhängig vom Lastgradient \dot{M}

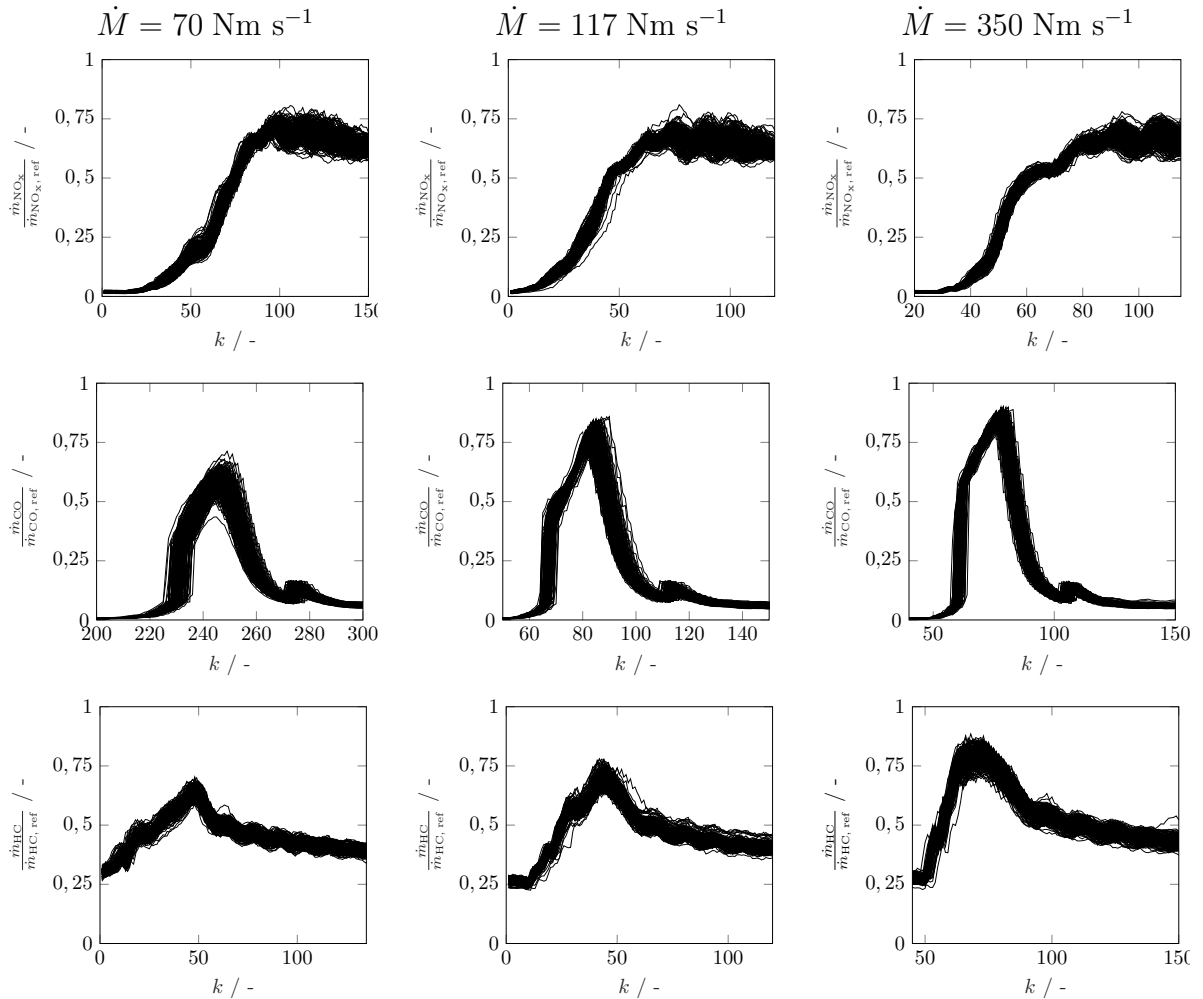


Abbildung A.3 Kurvenschar der Emission-Massenströme \dot{m}_{NO_x} , \dot{m}_{CO} und \dot{m}_{HC} über die Laständerung abhängig vom Lastgradient \dot{M}

Die Ergebnisse in Tab. A.1 stellen die prozentualen Ergebnisse der Optimierung der statischen Modellierungen relativ zur Ursprungsapplikation dar. Der Vertrauensbereich γ entspricht dem 95 %-Konfidenzintervall der Modelle. Die Modellprädiktionen und Messungen von Validierungsdaten, Wiederholpunkten und Messwerten sind für die integralen HC-Werte in Abb. A.5 dargestellt.

Tabelle A.1 Ergebnisse der Optimierung der statischen Modelle und $\gamma_{95\%}$ Vertrauensbereich der Modelle

\hat{y}	\dot{M}	$\gamma_{95\%}$	Optimierung
NO _x	350 Nm s ⁻¹	±3%	99%
NO _x	117 Nm s ⁻¹	±6%	90%
NO _x	70 Nm s ⁻¹	±6%	87%
HC	350 Nm s ⁻¹	±2%	98%
HC	117 Nm s ⁻¹	±3%	101%
HC	70 Nm s ⁻¹	±3%	99%
CO	350 Nm s ⁻¹	±3%	103%
CO	117 Nm s ⁻¹	±4%	93%
CO	70 Nm s ⁻¹	±3%	84%

Die nichtlinearen Optimierungsergebnisse der Zielgrößen sind in Abb. A.4 dargestellt. In Abb. A.4a sind die zweidimensionalen Pareto-Fronten der statischen Optimierung für die integralen Skalare der NO_x und CO-Kriterien gezeigt. Der iterative Suchpfad der Zielfunktion einer PSO der nichtlinearen dynamischen Modellierung für den Lastgradient $\dot{M} = 350 \text{ Nm s}^{-1}$ ist in Abb. A.4b dargestellt. Der Zielfunktionswert ergibt sich aus einer gleichgewichteten Summierung der Emissionsbestandteile entlang der positiven Laständerung mit der Beziehung

$$z_{\text{ZF}} = \int \dot{m}_{\text{NO}_x} dt + \frac{\overline{\dot{m}_{\text{NO}_x}}}{\overline{\dot{m}_{\text{HC}}}} \int \dot{m}_{\text{HC}} dt. \quad (\text{A.1})$$

Die Messwerte der integralen Kriterien sind über deren Variation sowie der Größe nach geordnet in den Abb. A.6 bis A.8 für die drei Emissionskomponenten NO_x, HC und CO gezeigt. Die Messausreißer sind ebenfalls dargestellt. Die visuellen Erklärbarkeitsansätze ICE und PDP der statischen Modelle in Abb. A.9 und Abb. A.10 der integralen Skalare zeigen den Einfluss der variierten Stützstelle u_1 bis u_2 auf die NO_x-, CO- und HC-Emission abhängig vom Lastgradient \dot{M} . Für den Lastgradient $\dot{M} = 350 \text{ Nm s}^{-1}$ sind die Modellauswertungen ICE und PCP für NO_x und CO in Abb. A.11 gezeigt. Die Modelle der HC-Kriterien sind aufgrund der mäßigen Modellqualität nicht aufgeführt.

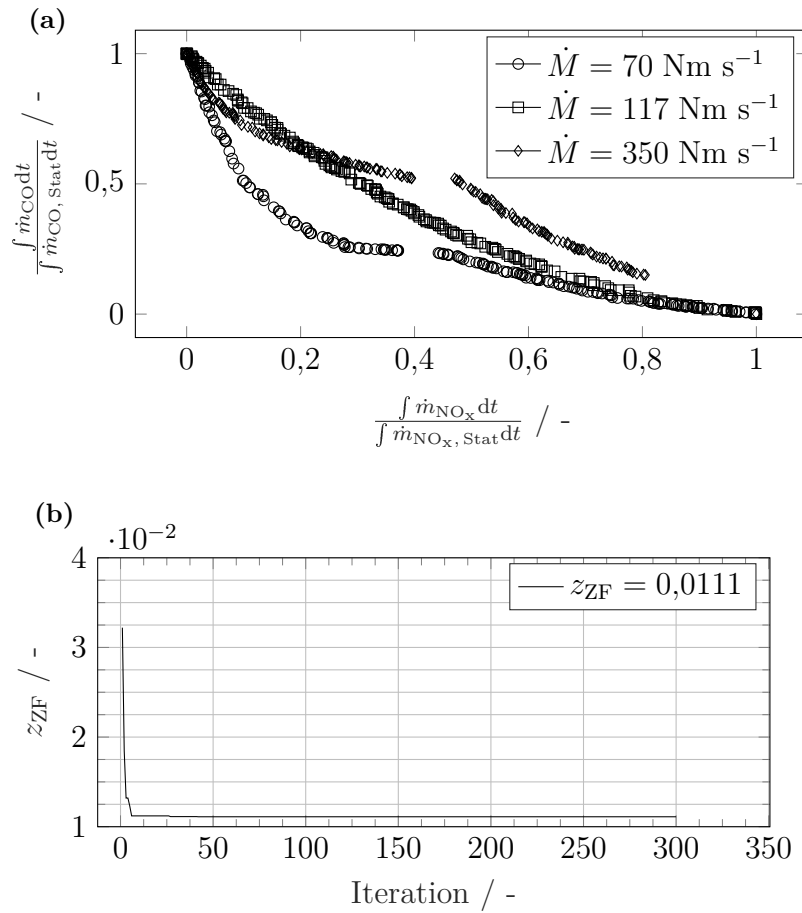


Abbildung A.4 Vergleich der zweidimensionalen Pareto-Fronten der statischen Optimierung mittels GMO (a) und die Zielfunktionsauswertungen der PSO für die Optimierung der nichtlinearen dynamischen Modellierung über 300 Iterationen

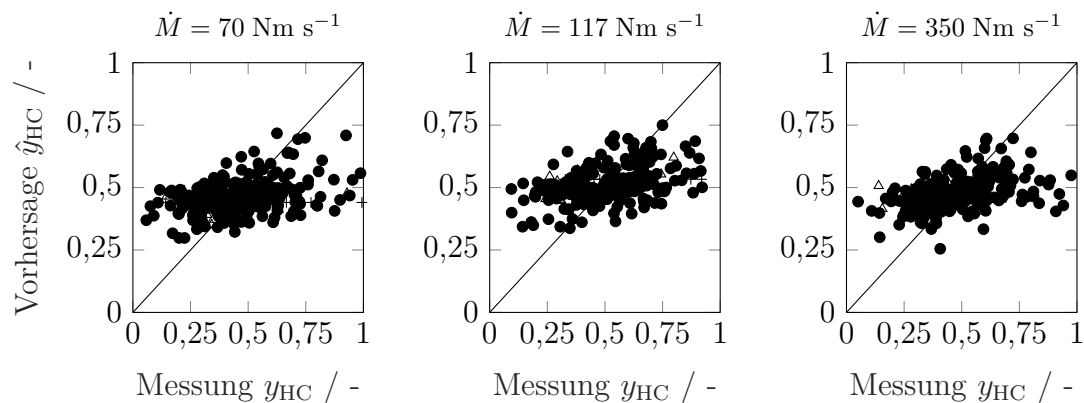


Abbildung A.5 Streudiagramme der Modellprädiktionen \hat{y} über den Messwerten y der skalaren HC-Größen in Abhängigkeit vom Lastgradient \dot{M} für Validierungsdaten (\triangle), Wiederholpunkte ($+$) und Messwerte (\bullet)

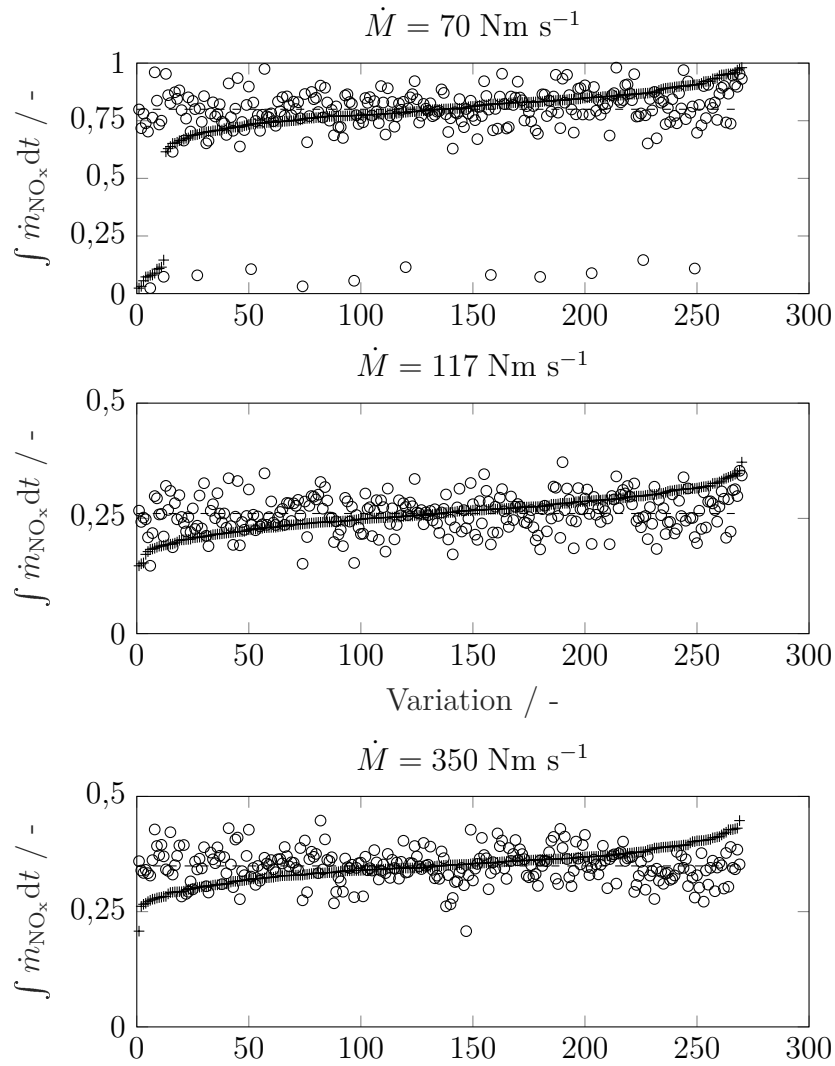


Abbildung A.6 Darstellung der integralen NO_x-Massenstromkriterien der statischen Modellierung inklusive der Messausreißer (○) über der Variation sowie der Größe nach geordnet (+)

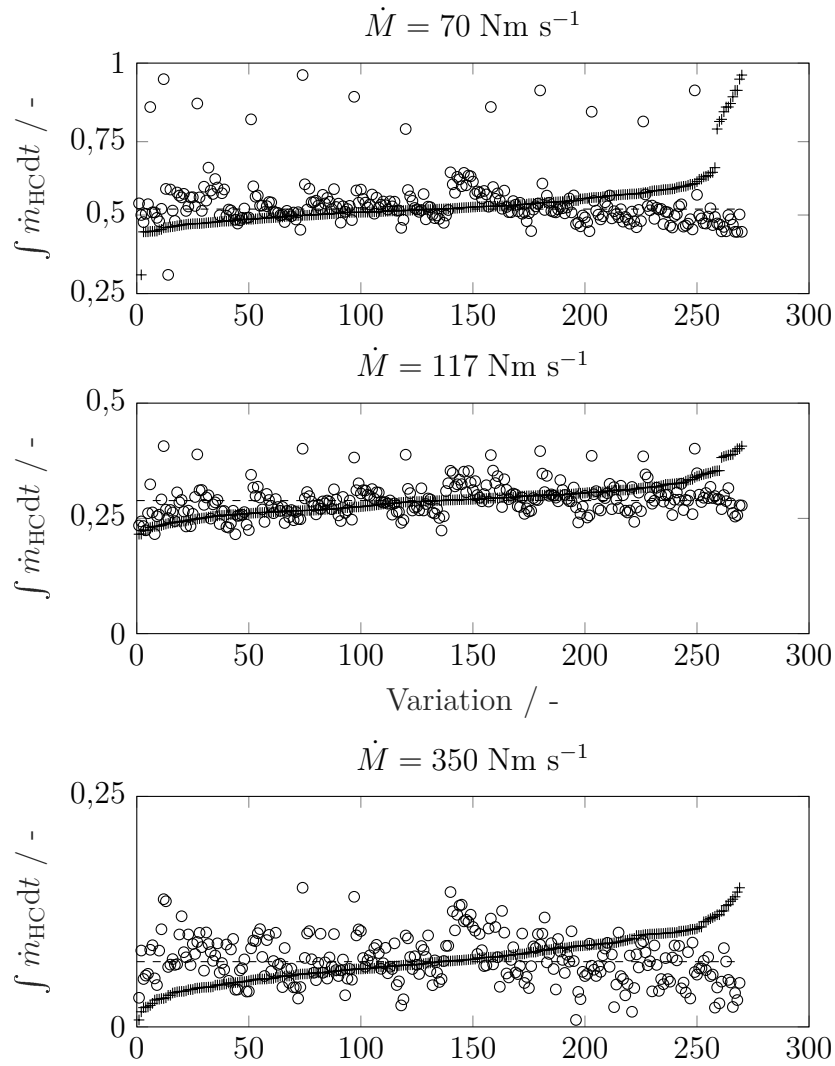


Abbildung A.7 Darstellung der integralen HC-Massenstromkriterien der statischen Modellierung inklusive der Messausreißer (\circ) über der Variation sowie der Größe nach geordnet ($+$)

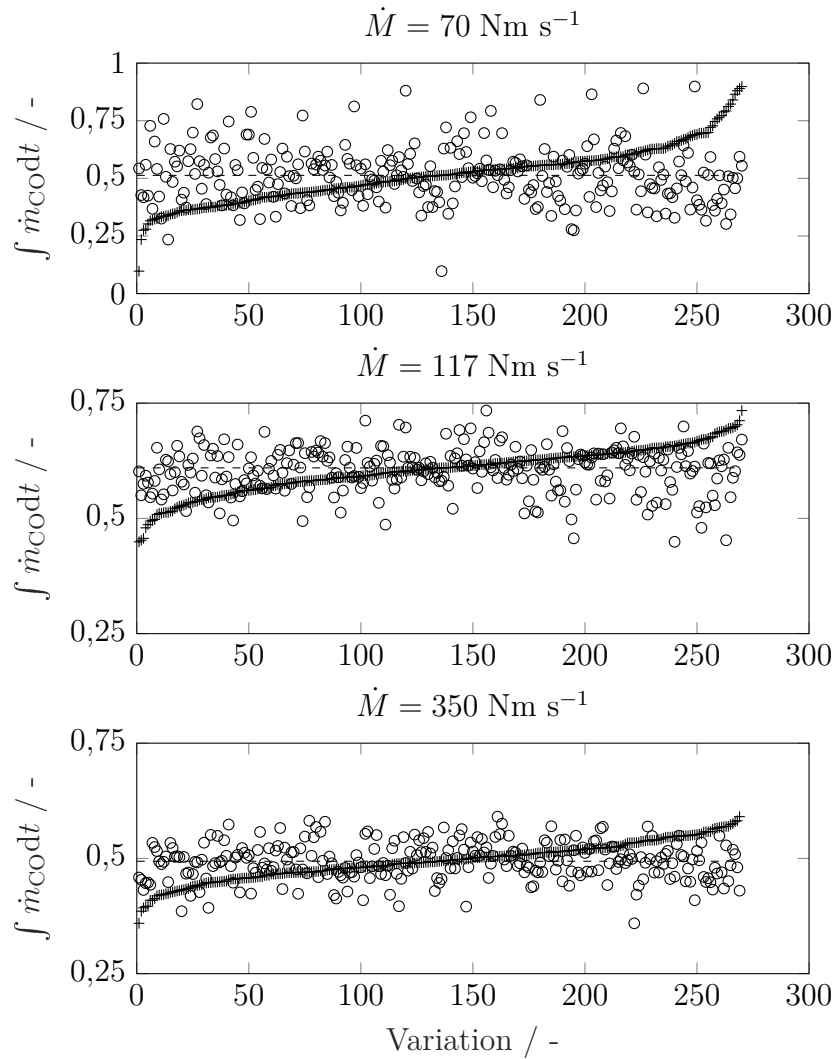


Abbildung A.8 Darstellung der integralen CO-Massenstromkriterien der statischen Modellierung inklusive der Messausreißer (\circ) über der Variation sowie der Größe nach geordnet ($+$)

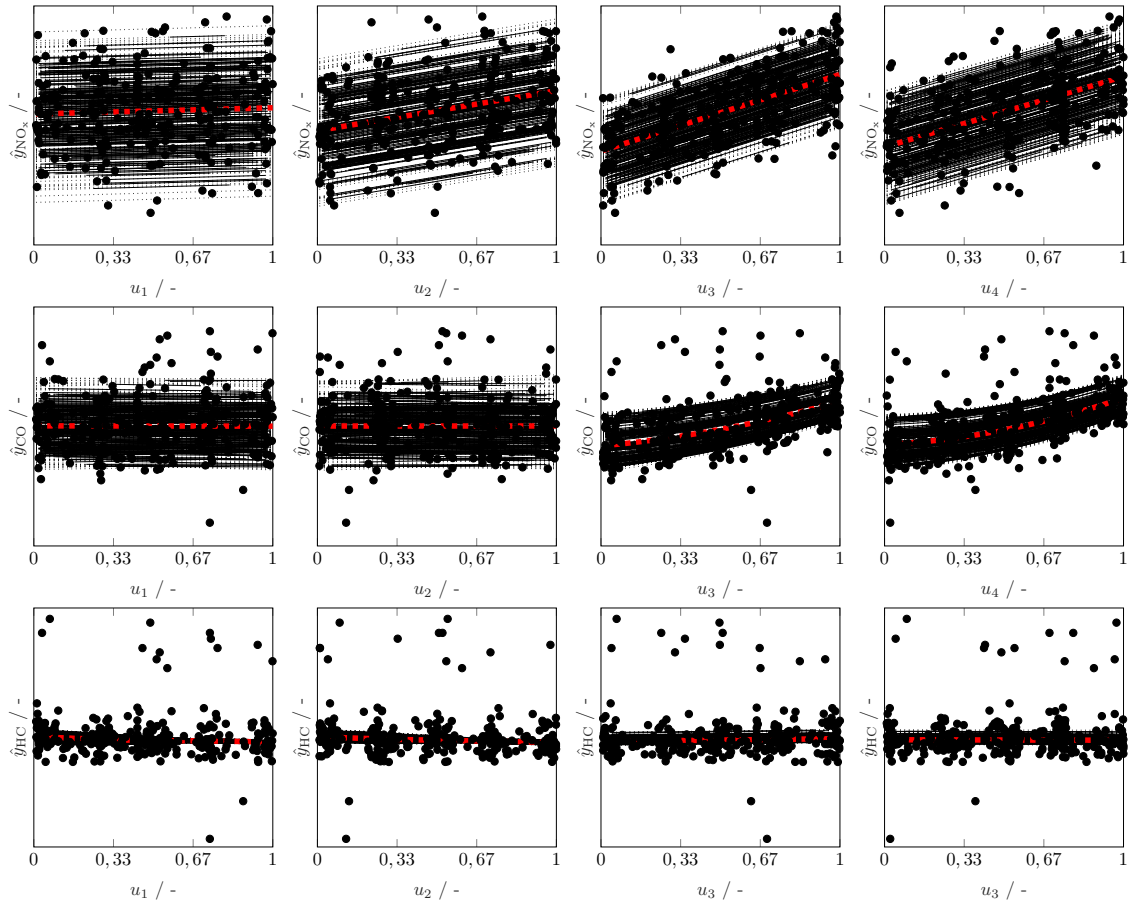


Abbildung A.9 PDP (rot) und ICE-Plot (schwarz) für die vier Verstellparameter u_1 bis u_4 des Einspritzzeitpunktes der Haupteinspritzung ϕ_{MI} im gesamten Faktori-Variationsbereich auf die Ausgangsgrößen \hat{y}_{NO_x} , \hat{y}_{CO} und \hat{y}_{HC} für den Lastgradient $\dot{M} = 70 \text{ Nm s}^{-1}$. Extrapolierte Versuchsraumverletzungen der Faktorkombinationen sind gepunktet, gültige Modellauswertungen als durchgezogene Linie und die PD dargestellt

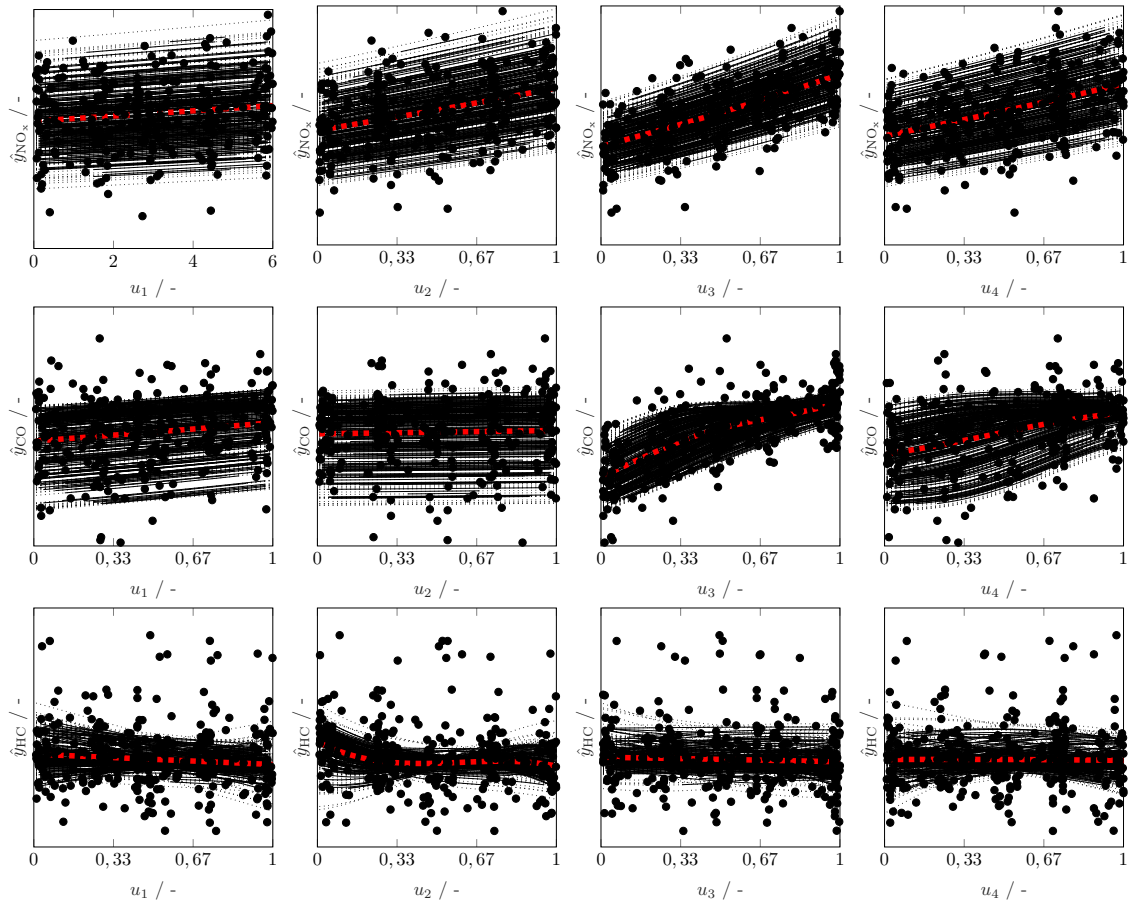


Abbildung A.10 PDP (rot) und ICE-Plot (schwarz) für die vier Verstellparameter u_1 bis u_4 des Einspritzzeitpunktes der Haupteinspritzung ϕ_{MI} im gesamten Faktori-Variationsbereich auf die Ausgangsgrößen \hat{y}_{NO_x} , \hat{y}_{CO} und \hat{y}_{HC} für den Lastgradient $\dot{M} = 117 \text{ Nm s}^{-1}$. Extrapolierte Versuchsraumverletzungen der Faktorkombinationen sind gepunktet, gültige Modellauswertungen als durchgezogene Linie und die PD dargestellt

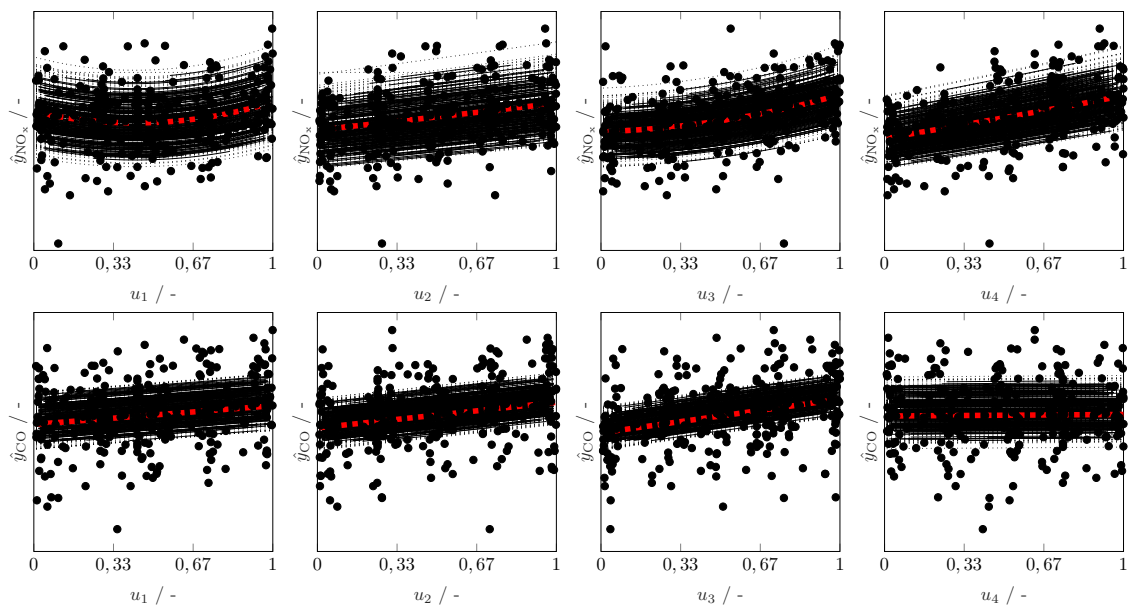


Abbildung A.11 PDP (rot) und ICE-Plot (schwarz) für die vier Verstellparameter u_1 bis u_4 des Einspritzzeitpunktes der Haupteinspritzung ϕ_{MI} im gesamten Faktoriationsbereich auf die Ausgangsgrößen \hat{y}_{NO_x} und \hat{y}_{CO} für den Lastgradient $\dot{M} = 350 \text{ Nm s}^{-1}$. Extrapolierte Versuchsraumverletzungen der Faktorkombinationen sind gepunktet, gültige Modellauswertungen als durchgezogene Linie und die PD dargestellt

B Grundlagen der Stochastik

B.1 Kenngrößen empirischer Verteilungen

Sowohl für die Berechnung der Korrelationskoeffizienten als auch für die Bestimmung der Sensitivitätsindices sind die Kenngrößen des Datensatzes bestimmen. Zur Beschreibung der Datensätze werden der Mittelwert [81, S. 65], als Streuungsparameter die empirische Varianz und die empirische Standardabweichung [81, S. 70] sowie als Zusammenhangsmaß einer zweidimensionalen Stichprobe, die empirische Kovarianz [81, S. 126] definiert.

Das arithmetische Mittel \bar{u} einer Wertmenge \mathbf{U} ist definiert als

$$\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i. \quad (\text{B.1})$$

Der Schätzwert für die empirische Varianz s_u^2 einer Wertmenge \mathbf{U}

$$s_u^2 = \frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2, \quad (\text{B.2})$$

und der Standardabweichung s_u sind definiert durch

$$s_u = \sqrt{s_u^2}. \quad (\text{B.3})$$

Der Schätzwert für die empirische Kovarianz zwischen zwei Wertmengen \mathbf{U} und \mathbf{Y} ist definiert als

$$s_{uy} = \frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})(y_i - \bar{y}). \quad (\text{B.4})$$

B.2 Fehlermaße

Zur Bewertung der Modellgüte werden folgende Gütemaße herangezogen, beziehungsweise erfolgt die Modellbewertung in dem DoE-Softwaretool *IAV EasyDoE* über hier vorgestellte Fehlermaße [55]. Diese beziehen sich auf die Messdatenpunkte y sowie die Modellvorhersagen \hat{y} . Die Division einer Quadratsumme mit der Anzahl

an Messungen N ergibt die mittlere quadratische Abweichung (engl. *Mean square error*, MSE)

$$\text{MSE} = \frac{1}{n} \sum_{k=1}^N (y_k - \hat{y}_k)^2, \quad (\text{B.5})$$

und deren Quadratwurzel (engl. *Root mean square error*, RMSE) zu

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{k=1}^N (y_k - \hat{y}_k)^2}. \quad (\text{B.6})$$

Eine Normierung der dimensionsbehafteten Fehlermaße gibt die normierte mittlere quadratische Abweichung (engl. *Normalized root mean square error*, NRMSE) mit dem Mittelwert \bar{y} zu [48, S. 76]

$$\text{NRMSE} = \sqrt{\frac{\sum_{k=1}^N (y_k - \hat{y}_k)^2}{\sum_{k=1}^N (y_k - \bar{y}_k)^2}}. \quad (\text{B.7})$$

Ein kleiner Wert des MSE, RMSE oder des NRMSE ergibt sich bei einer guten Anpassung des Modells. Zur Bewertung der nichtlinearen dynamischen Modellen wird das Bestimmtheitsmaß R^2 eingeführt. Das Bestimmtheitsmaß R^2 beschreibt die Güte der Modellierung der Beobachtungen durch das Modell. Dies ist definiert als Verhältnis der Modellvorhersage zur Messung, jeweils auf den Mittelwert bezogen und hängt folglich nur von den Residuen ab mit [115, S. 236]

$$R^2 = 1 - \frac{\sum_{k=1}^N (y_k - \hat{y}_k)^2}{\sum_{k=1}^N (y_k - \bar{y}_k)^2}. \quad (\text{B.8})$$

Das Bestimmtheitsmaß R^2 liegt im Bereich $[0, 1]$, wobei hohe Werte auf eine gute Anpassung des Modells hindeuten. Zur Bestimmung des schlechtesten lokalen Modells beim iterativen HILOMOT-Algorithmus wird der gewichtete quadratische Modellfehler [48, S. 76]

$$E_{i,\text{lokal}} = \sum_{k=1}^N \Phi_i(z_k) e^2(k), \quad (\text{B.9})$$

verwendet. Eine Überanpassung des Modells kann mit den vorgestellten Fehlermaßen nicht erfasst werden. Deshalb wird bei der statischen Modellbildung zudem für alle Messdaten ein Modell antrainiert, wobei ein Gütemaß verwendet wird, welches einen nicht berücksichtigten Messpunkt aus der Parameterschätzung herausrechnet und auf diesen einen quadratischen Fehler berechnet. Es wird folglich ein Fehler eines Versuchspunktes berechnet, ohne den betrachteten Punkt in der Schätzung der Modellkoeffizienten zu verwenden. Die Quadratsumme der Vorhersagefehler (engl. *Predicted sum of squares*, PRESS) berechnet sich mit [69, S. 35, 91, S. 43, 82, S. 418, 30, S. 27, 106]

$$\text{PRESS} = \sum_{k=1}^N (y_k - \hat{y}_{(k)})^2. \quad (\text{B.10})$$

Für diese Vorhersagen verwendet das DoE-Softwaretool *IAV EasyDoE* eine Betrachtung des NRMSE an, die zudem von der Art des Modells abhängt, mit [55]

$$\text{NRMSE}_{\text{PRESS}} = \sqrt{\frac{\sum_{k=1}^N (y_k - \hat{y}_{(k)})^2}{\sum_{k=1}^N (y_k - \bar{y}_k)^2}}. \quad (\text{B.11})$$

B.3 Rangtransformation

Bei der Rangtransformation einer Zeitreihe $u(k)$ zu den Zeitpunkten $k = [0, 1, \dots, N]$ beschreibt der Rang die Transformation

$$r : u(k) \rightarrow R(k). \quad (\text{B.12})$$

Der kleinste Wert u_{\min} der Zeitreihe bekommt den Rang $R(u_{k, \min}) = 1$, der zweitkleinste Wert den Rang $R(u_{k, \min}) = 2$ gesetzt und jeder weitere Wert wird nach dem jeweiligen Rang sortiert. Dies ist genau dann eindeutig umkehrbar, wenn alle Werte der Zeitreihe verschieden sind. Die relative Rangfolge $r(k) = R(k)/N$ ist auf dem Intervall $[0, 1]$ gleichverteilt für eine Zeitreihe mit ausschließlich unterschiedlichen Werten. Liegen für die Zeitreihe identische numerische Werte u mehrfach an verschiedenen Stellen k in der Datenmenge vor, gelten für diese identischen Werte u_{ident} der Mittelwert

$$R(u_{\text{ident}}) = \frac{R(u_{\text{ident, gr}}) + R(u_{\text{ident, kl}})}{2}. \quad (\text{B.13})$$

Es entspricht $R(u_{\text{ident, gr}})$ dem Rang des ersten Wertes größer als u_{ident} und $R(u_{\text{ident, kl}})$ dem ersten Wert kleiner als u_{ident} für die mehrfachen identischen numerischen Werte u_{ident} der Zeitreihe. Liegen paarweise identische numerische Werte vor, ist eine Gleichverteilung der Rangfolge nicht mehr hinreichend gegeben. Eine Gleichverteilung in der Häufigkeitsverteilung der Datenreihe kann durch eine Rangtransformation erreicht werden, bei der die Zeitreihe aus unterschiedlichen numerischen Werten besteht. Abhilfe schafft die Transformation von [38]. Die Zeitreihe $u(k)$ mit den o identischen Punkten $u_{\text{ident}} = [u(k_{(1, \text{ident})}), u(k_{(2, \text{ident})}), \dots, u(k_{(o, \text{ident})})]$ mit dem Wert K wird mit

$$\begin{aligned} s : u_{\text{ident}} &\rightarrow v(k) = K + a_o; \\ \text{mit } 0 &\leq a_1 \leq a_2 \leq \dots \leq a_o \leq 1 \end{aligned} \quad (\text{B.14})$$

zur Zeitreihe $v(k)$ abgebildet. In der Zeitreihe $v(k)$ sind sämtliche Werte mit $k = [0, 1, \dots, N]$ paarweise verschieden. Eine Abbildungsfolge

$$t : u(k) \xrightarrow{s} v(k) \xrightarrow{r} R(k) \quad (\text{B.15})$$

kann mit der Gl. B.12 und der Gl. B.14 kombiniert werden. Mit anschließender Normierung ist eine relative Rangfolge $r(k) = R(k)/N$ im Intervall $[0, 1]$ gleichverteilt.

Mit der Transformation können beliebig verteilte endliche Zeitreihen in gleichverteilte Rangfolgen umgewandelt werden. Die Charakteristik der Zeitreihen bleibt im Wesentlichen bestehen. Ein Beispiel veranschaulicht das Vorgehen der Rangtransformation in Tab. B.1 detaillierter.

Tabelle B.1 Vergleich der Rangtransformationen r und t in Anwendung auf die Zeitreihe $u(k)$ zu den Zeitpunkten $k = [1, \dots, 9]$

k	1	2	3	4	5	6	7	8	9
$u(k)$	1	3	5	1	6	6	90	4	5.5
$r : u(k) \xrightarrow{r} R(k)$	1,5	3	5	1,5	7,5	7,5	9	4	6
$t : u(k) \xrightarrow{s} v(k) \xrightarrow{r} R(k)$	1	2	4	0	6	7	8	3	5

In Tab. B.1 ist eine beliebige Zeitreihe $u(k)$ mit $k = 9$ Abtastschritten im Intervall $[1, 90]$ gegeben. Die Zeitreihe weist den mehrfach identischen Zahlenwert $u_{\text{ident}} = 1$ sowie den mehrfach den identischen Zahlenwert $u_{\text{ident}} = 6$ auf. Das Beispiel soll die unterschiedlichen Ergebnisse der Rangtransformationen r und t anhand der Beispieldaten verdeutlichen. Bei der Transformation r nach Gl. B.12 werden die Ränge der identischen Zahlenwerte der Zeitreihe mit der Gl. B.13 gemittelt. Im ersten Fall ergibt sich der Mittelwert für das Rangpaar $R(u_{\text{ident,gr}}) = 3$ und $R(u_{\text{ident,kl}}) = 0$ zu $R(u_{\text{ident}}) = 1,5$. Für den Fall $u_{\text{ident}} = 6$ ergibt sich der Rang analog mit dem Rangpaar $R(u_{\text{ident,gr}}) = 9$ und $R(u_{\text{ident,kl}}) = 6$ der Rang $R(u_{\text{ident}}) = 7,5$. Entsprechend dieser Logik folgt die empirische Schätzung des Rangkorrelationskoeffizient nach *Spearman*. Eine Häufigkeitsverteilung der rangtransformierten Zeitreihe $R(k)$ bildet in diesem Fall jeweils Häufungen an den Stellen der mehrfach identischen Zahlenwerte. Zusätzlich ist der Einfluss der Rangtransformation nützlich bei der Behandlung möglicher Messausreißer. Dies ist exemplarisch im Beispiel der Tab. B.1 für $u(k = 7) = 90$ gezeigt. Die Korrelationsanalyse nach *Pearson* kompensiert Ausreißer nicht.

Wird die Abbildungsfolge der Gl. B.15 verwendet, wird die Zeitreihe durch die vorgelagerten Transformation s der Gl. B.14 paarweise verschieden aufbereitet. Je nach Anzahl der identischen Punkten o wird zwischen dem ersten Zahlenwert größer und kleiner als u_{ident} ein Abstandswert a_i ermittelt. Der Abstandswert wird auf u_{ident} addiert und für jeden einzelnen mehrfach identischen Zahlenwert erhöht. Die resultierende Zeitreihe $v(k)$ weist folglich paarweise verschiedene Zahlenwerte auf. Wird diese wiederum mit der Transformation r nach Gl. B.12 umgewandelt, entspricht dies den Werten der Abbildung t der Gl. B.15. Eine Gleichverteilung tritt in diesem Beispiel ausschließlich in der Transformation t auf. Die Abbildungsfolge t der Gl. B.15 ist hier von entscheidendem Vorteil in der Schätzung der Transinformation in Kap. 3.1.2 und daher an dieser Stelle eingeführt.

B.4 Korrelationsintegral

Zur Schätzung der Transinformation ist der Ansatz vorgestellt nach [38, 94, 93]. Zur Bestimmung der Wahrscheinlichkeiten sind die Korrelationsintegrale aller Punktepaare zu bilden. Das Korrelationsintegral kann nach [94] wie folgt formuliert werden [115]

$$C(\varepsilon) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \{|U - U| < \varepsilon\}. \quad (\text{B.16})$$

Der Wert des Korrelationsintegral ist anhand prozentualer Erfahrungswerte für die Schätzung der Transinformation so einzustellen, dass ein geeigneter Wert für die Vergrößerung ε_d iterativ gefunden wird. Ein von der Dimension unabhängiges Maß ergibt sich dabei für einen kleinen, von Null verschiedenen, konstanten Wert des Korrelationsintegrals $C_{U \wedge Y, N}$, unabhängig von den gewählten Dimensionen d_1 und d_2 [38, S. 42]. Die Zielfunktion ist $H_2(\mathbf{U}, \mathbf{Y}, \varepsilon_d) = -\text{ld}(0.05)$ [38, S. 42]. Damit beträgt $C_{U \wedge Y, N} = 0,05$ und die Binärmatrix $\mathcal{B}_{U \wedge Y}$ ist zu 5% besetzt. Der iterative Suchpfad einer Innere-Punkte Optimierung und die Verlustfunktion ist in Abb. B.1 gezeigt.

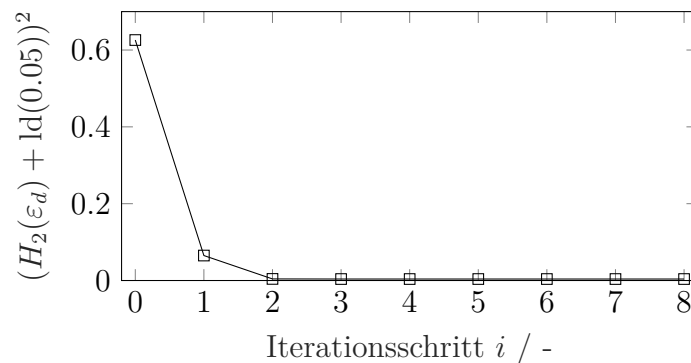


Abbildung B.1 Iterative Bestimmung von ε_d zur Lösung der Zielfunktion $H_2(\mathbf{U}, \mathbf{Y}, \varepsilon_d) = -\text{ld}(0.05)$ mittels nichtlinearer Innere-Punkte-Optimierung

B.5 Herleitung der Sensitivitätsanalyse nach Sobol

Die von Sobol [117] eingeführte Methode bietet eine varianzbasierte, globale Methode um Sensitivitätsmaße zu berechnen. Sobols Idee basiert auf der quadratisch integrierbaren Funktion f im n_p -dimensionalen Einheitshyperkubus

$$\Omega^{n_p} = (\mathbf{U} | 0 \leq u_i \leq 1, \quad i = 1, 2, \dots, n_p) \quad (\text{B.17})$$

mit den Eingangsparametervektor $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_p}]^\top$. Die Funktion f kann in Form einer *High-dimensional model representation* (HDMR) in einzelne eingangspa-

parameterabhängige Summanden zerlegt werden. Die HDMR einer 2^{n_p} maximalen Anzahl an Summanden

$$Y = f(\mathbf{u}) = f_0 + \sum_{i=1}^{n_p} f_i(u_i) + \sum_{i=1}^{n_p} \sum_{j=i+1}^{n_p} f_{i,j}(u_i, u_j) + \dots + f_{1,2,\dots,n_p}(u_1, u_2, \dots, u_{n_p}), \quad (\text{B.18})$$

ist in der jeder Term im Eingangsraum quadratisch integrierbar [102]. Die Zerlegung ist dabei endlich. Der Term nullter Ordnung ist f_0 . Die n_p -Terme von f_i sind die der ersten Ordnung. Die n_p -Terme von $f_{i,j}$ sind die der zweiten Ordnung und die Terme $f_{1,2,\dots,n_p}$ die der n_p -ten Ordnung. Der Ausdruck aus Gl. B.18 besitzt die Gültigkeit wenn die Bedingung

$$\int_0^1 f_{i_1 \dots i_s}(u_{i_1}, \dots, u_{i_s}) du_{i_{n_p}} = 1, \quad \text{mit } n_p = [1, \dots, s] \quad (\text{B.19})$$

erfüllt ist [117]. Für den Fall, dass die Mittelwerte eines jeden einzelnen Ausdrucks aus Gl. B.18 null betragen, sind die Terme nach Sobol orthogonal zueinander [117]. Folglich lassen sich alle Terme in Gl. B.18 durch mehrdimensionale Integrale eindeutig bestimmen [102, S. 160, 20, S. 4]. Der Erwartungswert des konstanten Terms f_0 als auch die bedingten Erwartungswerte der Modellausgänge \mathbf{Y} sind gegeben nach [102] durch

$$f_0 = \mathbb{E}\{\mathbf{Y}\}, \quad (\text{B.20})$$

$$f_i(u_i) = \mathbb{E}\{\mathbf{Y}|u_i\} - \mathbb{E}\{\mathbf{Y}\}, \quad (\text{B.21})$$

$$f_{i,j}(u_i, u_j) = \mathbb{E}\{\mathbf{Y}|u_i, u_j\} - f_i(u_i) - f_j(u_j) - \mathbb{E}\{\mathbf{Y}\}. \quad (\text{B.22})$$

Die Varianzen der Terme, in Abhängigkeit vom bedingten Erwartungswert, bezeichnen das Maß der Sensitivität. Die Varianz von beispielsweise Gl. B.21 $\sigma_{f_i(u_i)}^2$ ist $\sigma_{Y|u_i}^2$. Wird diese bedingte Varianz ins Verhältnis zur Gesamtvarianz σ_Y^2 gesetzt, ergibt sich der Sensitivitätsindex der ersten Ordnung Gl. 3.14. Der Index beschreibt die Haupteffekte eines Eingangsparameters auf die Ausgangsgröße. Interaktionseffekte von u_i and u_j werden mit den Varianzen der Verbundeffekte Gl. B.22 durch [102, S. 161]

$$\sigma_{i,j}^2 = \sigma_{f_{i,j}(u_i, u_j)}^2 = \sigma_{Y|u_i, u_j}^2 - \sigma_{Y|u_i}^2 - \sigma_{Y|u_j}^2, \quad (\text{B.23})$$

berechnet. Den Verbundeffekt des Eingangspaars (u_i, u_j) auf den Modellausgang bezeichnet $\sigma_{f_{i,j}(u_i, u_j)}^2$. Der Varianzterm $\sigma_{i,j}^2$ bildet die gemeinsamen Effekte der Parameter minus die Haupteffekte von u_i und u_j . Der Term $\sigma_{i,j}^2$ wird als Effekt der zweiten Ordnung bezeichnet und kann auf beliebig endliche höhere Ordnungen im Eingangsraum Ω^{n_p} erweitert werden [102, S. 162]. Die sog. *Functional analysis of variance* (FANOVA)

$$\sigma_Y^2 = \sum_{i=1}^N \sigma_i^2 + \sum_{i=1}^N \sum_{j=i}^M \sigma_{i,j}^2 + \dots + \sigma_{1,2,\dots,n_p}^2 \quad (\text{B.24})$$

bildet die Expansion der Terme bis zur Ordnung n_p ab [102, S. 162]. Die Sensitivitätsindizes sämtlicher Haupt- und Interaktionseffekte ergeben sich durch eine Normierung der einzelnen Terme mit der Gesamtvarianz σ_Y^2 [102, S. 161]

$$\sum_{i=1}^N S_i + \sum_{i=1}^N \sum_{j=i}^M S_{i,j} + \sum_{i=1}^N \sum_{j=i}^M \sum_{k=j}^L S_{i,j,k} + \dots = 1. \quad (\text{B.25})$$

Dabei bezeichnen die Effekte der ersten Ordnung S_i , die der zweiten Ordnung $S_{i,j}$ sowie die der dritten Ordnung $S_{i,j,k}$. Der modellunabhängige Totaleffekt ist definiert als Ausgangsvarianz sämtlicher Effekte, in Abhängigkeit von einem Eingangsfaktor und gilt auch für nicht-additive Modelle [102, S. 163].

B.5.1 Monte-Carlo-Schätzung der varianzbasierten Sensitivitätsindizes

Die Varianzen der Terme mit bedingten Erwartungswerten aus Gl. B.20 - B.22 und die Sensitivitätsindizes aus Gl. 3.16-3.17 werden nach einer Generierung von Stichproben basierend auf einer Verteilungsdichte Funktion mittels numerischer Quasi-Monte-Carlo-Methode berechnet [102, S. 164ff]. Der Vorgang wird hier zusammengefasst nach [102, S. 164ff] eingeführt wie folgt:

1. Bilden zweier Matrizen \mathbf{A} und \mathbf{B} der Dimension $N \times 2P$ aus Zufallszahlen, mit der Anzahl an Eingängen P und dem Stichprobenumfang N als

$$\mathbf{A} = \begin{bmatrix} u_1^{(1)} & u_2^{(1)} & \dots & u_i^{(1)} & \dots & u_P^{(1)} \\ u_1^{(2)} & u_2^{(2)} & \dots & u_i^{(2)} & \dots & u_P^{(2)} \\ \vdots & \vdots & & \vdots & & \vdots \\ u_1^{(N)} & u_2^{(N)} & \dots & u_i^{(N)} & \dots & u_P^{(N)} \end{bmatrix}, \quad (\text{B.26})$$

$$\mathbf{B} = \begin{bmatrix} u_{P+1}^{(1)} & u_{P+2}^{(1)} & \dots & u_{P+i}^{(1)} & \dots & u_{2P}^{(1)} \\ u_{P+1}^{(2)} & u_{P+2}^{(2)} & \dots & u_{P+i}^{(2)} & \dots & u_{2P}^{(2)} \\ \vdots & \vdots & & \vdots & & \vdots \\ u_{P+1}^{(N)} & u_{P+2}^{(N)} & \dots & u_{P+i}^{(N)} & \dots & u_{2P}^{(N)} \end{bmatrix}. \quad (\text{B.27})$$

2. Definition einer Matrix \mathbf{C} gleich der Matrix \mathbf{B} ohne der i -ten Spalte, welche von Matrix \mathbf{A} genommen wird, als

$$\mathbf{C}_i = \begin{bmatrix} u_{P+1}^{(1)} & u_{P+2}^{(1)} & \dots & u_i^{(1)} & \dots & u_{2P}^{(1)} \\ u_{P+1}^{(2)} & u_{P+2}^{(2)} & \dots & u_i^{(2)} & \dots & u_{2P}^{(2)} \\ \vdots & \vdots & & \vdots & & \vdots \\ u_{P+1}^{(N)} & u_{P+2}^{(N)} & \dots & u_i^{(N)} & \dots & u_{2P}^{(N)} \end{bmatrix}. \quad (\text{B.28})$$

3. Berechnung der Modellausgänge in Form von Vektoren der Länge $N \times 1$ für jede Matrix \mathbf{A} , \mathbf{B} und \mathbf{C}_i

$$\mathbf{y}_A = f(\mathbf{A}) \quad , \quad \mathbf{y}_B = f(\mathbf{B}) \quad \text{und} \quad \mathbf{y}_{C_i} = f(\mathbf{C}_i). \quad (\text{B.29})$$

4. Bestimmung der Haupteffekte nach Gl. 3.16 mit Hilfe des Skalarprodukts aus den Vektoren zu

$$S_i = \frac{\sigma_{\mathbb{E}\{Y|u_i\}}^2}{\sigma_Y^2} = \frac{\mathbf{y}_A \cdot \mathbf{y}_{C_i} - f_0^2}{\mathbf{y}_A \cdot \mathbf{y}_A - f_0^2}, \quad (\text{B.30})$$

$$\text{mit} \quad f_0^2 = \left(\frac{1}{N} \sum_{j=1}^N \mathbf{y}_A^{(j)} \right)^2. \quad (\text{B.31})$$

5. Berechnung der Totaleffekte nach Gl. 3.17 mit

$$S_{T,i} = 1 - \frac{\sigma_{\mathbb{E}\{Y|u_{\sim i}\}}^2}{\sigma_Y^2} = \frac{\mathbf{y}_B \cdot \mathbf{y}_{C_i} - f_0^2}{\mathbf{y}_A \cdot \mathbf{y}_A - f_0^2}. \quad (\text{B.32})$$

B.5.2 Fourier Amplitude Sensitivity Test

Eine Sensitivitätsanalyse im Frequenzbereich bietet der *Fourier amplitude sensitivity test* (FAST) [25, nach]. Ein Vorteil zur Methode nach Sobol ist der geringere rechnerische Aufwand, der speziell bei höherdimensionalen Funktionen durch die vielen Simulationsdurchläufe der Monte-Carlo-Methode entsteht. Die FAST Methode betrachtet eine eindimensionale Funktion. Für die Transformation wird jeder Eingangsgröße u_i eine Frequenz ω zugeordnet [102, S. 167] mit:

$$u_i(s) = g(\sin \omega_i s), \quad \forall i = 1, 2, \dots, n_p \quad (\text{B.33})$$

$$\text{mit} \quad g(s) = \frac{1}{2} + \frac{1}{\pi} \omega_i s, \quad (\text{B.34})$$

Die Funktion $g(s)$ wird abhängig von der Wahrscheinlichkeitsfunktion der Eingangsgröße u_i gewählt und der Parameter s variiert im Bereich $[-\pi, \pi]$. Die Eingangsparameter eines jeden Faktors werden mit den Frequenzen $\boldsymbol{\omega} = [\omega_1, \omega_2, \dots, \omega_{n_p}]$ über die Länge N abgetastet. Die Funktion $g(s)$ folgt [20]. Linear unabhängige Frequenzen $\boldsymbol{\omega}$ schlägt [25] vor. Die Funktion Gl. B.34 deckt den gesamten normierten eindimensionalen Eingangsbereich ab. Der Erwartungswert des Ausgangs der Funktion $f(s)$

$$\mathbb{E}\{\mathbf{Y}\} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) ds, \quad (\text{B.35})$$

$$\text{mit} \quad f(s) = f(g(\sin(\omega_1 s)), g(\sin(\omega_2 s)), \dots, g(\sin(\omega_{n_p} s))),$$

ist in Abhängigkeit von $g(s)$ definiert [113, S. 35]. Den Erwartungswert Gl. B.35 und die Gesamtvarianz σ_Y^2 folgt aus einer Fourier-Reihenentwicklung der Funktion $f(s)$ mit den Fourier-Koeffizienten A_k , B_k und A_0

$$\begin{aligned} f(s) &= A_0 + \sum_{k=1}^{\infty} (A_k \cos(ks) + B_k \sin(ks)), \\ A_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(s) \cos(ks) ds, \quad \forall k = 1, 2, \dots, \infty, \\ B_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(s) \sin(ks) ds, \\ A_0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) ds. \end{aligned} \tag{B.36}$$

Der Schätzwert der Varianz des Modellausgangs σ_Y^2 ergibt sich aus den Fourier-Koeffizienten der Reihenentwicklung Gl. B.36 mit [25]

$$\sigma_Y^2 \approx \frac{1}{2} \sum_{k=1}^{\infty} (A_k^2 + B_k^2). \tag{B.37}$$

Der Einfluss der Größe u_i auf die Ausgangsvarianz σ_Y^2 ergibt sich aus der zum Eingang zugehörigen Frequenz ω_i sowie der ganzzahligen vielfachen Oberschwingung. Die Oberschwingungen werden wegen des Alias-Effekts bis zur Ordnung n_m berücksichtigt, da mit zunehmender Ordnung die Oberschwingungen numerisch an Bedeutung verlieren [113, S. 36]. Die bedingte Varianz $\sigma_{Y|u_i}^2$ des Eingangs u_i berechnet sich mit [113, S. 36]

$$\sigma_{Y|u_i}^2 = \frac{1}{2} \sum_{p=1}^{n_m} (A_{p\omega_i}^2 + B_{p\omega_i}^2). \tag{B.38}$$

Der Fehler, der durch eine Begrenzung der Ordnung n_m entsteht, ist zu vernachlässigen. Cukier u. a. empfiehlt eine Ordnung $n_m = 4$ oder $n_m = 6$ [25, S. 9]. Eine höhere Ordnung erweitert den Stichprobenumfang zu groß. Das Verhältnis der i -ten Eingangsvarianz zur Ausgangsvarianz ergibt den FAST Sensitivitätsindex [102]

$$S_{\text{FAST}, i} = \frac{\sigma_{Y|u_i}^2}{\sigma_Y^2}. \tag{B.39}$$

C Weiterführende Gleichungen zum Stand der Technik

C.1 Statistische Versuchsplanung

Ein konventioneller Ansatz zur Vermessung eines Parameterraums kann mit vollfaktorieller Rastervermessung erfolgen. Der Versuchsaufwand, also die Anzahl der Punkte einer Rastervermessung n , berechnet sich für eine Anzahl an Stufen oder Level l in jede Eingangsdimension d , bzw. Anzahl Eingangsgrößen, zu [69, S. 8]

$$n = l^d. \quad (\text{C.1})$$

In der Versuchsplanung ist eine Rastervermessung unvorteilhaft, da bei steigender Anzahl zu variierender Kalibrierparameter der Messumfang exponentiell steigt. Eine erhebliche Reduzierung des Versuchsaufwandes kann durch eine statistische Versuchsplanung erreicht werden und ist daher eine weitverbreitete Methodik in der modellbasierten Kalibrierung. Die Anzahl Parameter und die Anzahl an Basisfunktionen eines Versuchsplans berechnen sich mit

$$n_M = \frac{(d + n_p)!}{d! n_p!} - 1 \quad (\text{C.2})$$

für ein n_p -dimensionales Polynommodell der Dimension d [85, S. 221, 69, S. 8, 48, S. 9]. Die Gleichung (Gl. C.2) beschreibt sowohl die Anzahl an Parametern $n_M + 1$, bzw. die Anzahl an Basisfunktionen n_M eines Polynommodells. Die Reduzierung des Versuchsaufwands durch die Nutzung einer Polynommodellstruktur im Vergleich zur Berechnung des Versuchsaufwandes mit einem vollfaktoriellen Versuchsplan ist in Abb. C.1 visualisiert. Trotz eines 50 % Punktüberhangs als Sicherheit gegenüber Messausreißern ist der Versuchsaufwand im Vergleich mit dem vollfaktoriellen Versuchsplan in Abb. C.1a durch eine modellbasierte Kalibrierung in Abb. C.1b erheblich verringert. Ein statistischer Versuchsplan nutzt Algorithmen um die Kandidatenpunkte nach bestimmten Kriterien im Versuchsraum auszuwählen.

Grundsätzlich sind drei verschiedene Typen von Versuchsplänen klassifizierbar. Hier werden zwei offline Methoden vorgestellt. Bei **optimalen Versuchsplänen** erfolgt die Versuchsplanung im Allgemeinen bei der Annahme einer linear parametrisierten Modellstruktur mit varianzminimierenden Designs [69, S. 47, 48, S. 97]. Ist aufgrund von zu wenig Information über das Prozessverhalten eine Vorselektion der

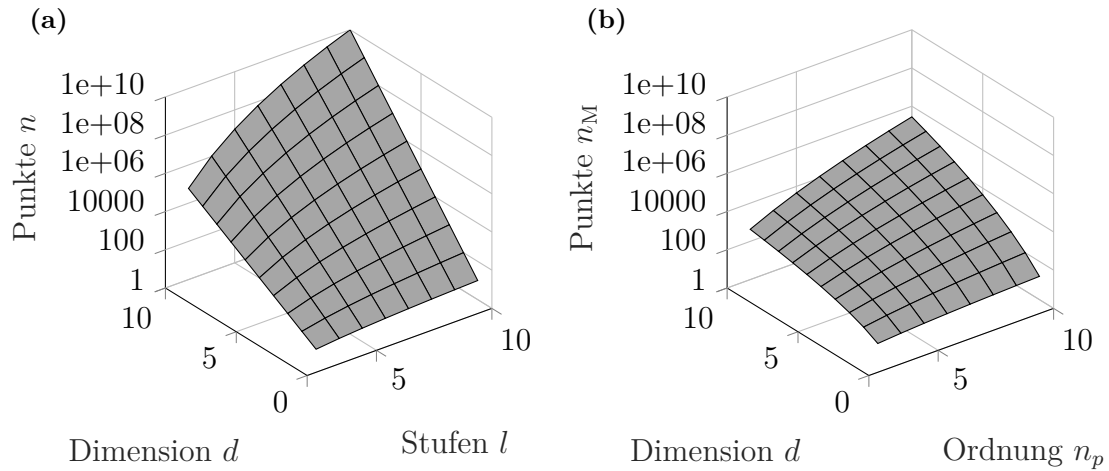


Abbildung C.1 Aufwand der Applikation bei der Versuchsplanung mit einem vollfaktoriellen Ansatz (a) oder einem modellbasierten Ansatz (b) mit einem 50% Punktüberhang [69, S. 7ff]

Modellstruktur nicht möglich, werden **raumfüllende Versuchspläne** eingesetzt. Die Kandidatenpunkte werden im gesamten Versuchsraum nach bestimmten Kriterien möglichst gleichmäßig verteilt. Zufällig im Raum verteilte Versuchspunkte gibt der Algorithmus *Latin hypercube sampling* aus [79]. Jede Stufe der Eingangsparameter wird in eine Anzahl an gleichen Randverteilungen eingeteilt und anschließend zufällig aus jeder dieser Schichten Kandidatenpunkte entnommen. Der Algorithmus erzeugt keine Gleichverteilung und folglich einen unkorrelierten Versuchsraum [35, S. 49, 79, 115, S. 206]. Die Versuchsplanung mit raumfüllenden Plänen empfiehlt sich bei Unstetigkeit im Systemverhalten, erhöht jedoch die Anzahl an benötigten Testpunkten für komplexere Modellstrukturen [115, S. 225]. Zusätzlich zu den Kandidatenpunkten werden Wiederholpunkte geplant, die unter anderem die Schätzung der Messvarianzen in diesen Anhäufungen der Kandidatenpunkte im Versuchsraum erleichtern und eine zeitliche Veränderung der Messdaten über die Versuchsdauer erfassen [59, S. 80].

C.2 Anpassung der Hyperparameter des Gaußprozess Modells

Die Hyperparameter eines GP werden mit einer *Marginal likelihood* (ML)-Methode bestimmt. Der Vektor Θ fasst alle Hyperparameter der Kernel-Funktion Gl. 2.7 zusammen $\Theta = \{\sigma_f^2, \ell_1, \dots, \ell_M, \sigma_n^2\}$. Das Training der Hyperparameter erfolgt über die Randwahrscheinlichkeit unter der Voraussetzung einer Übereinstimmung der multivariaten Normalverteilung und den Trainingsdaten. Die logarithmische ML-Funktion für die Randverteilung ist definiert als [115, S. 300, 60, S. 187, 96, S. 29]

$$\log p(\mathbf{y}|\mathbf{U}, \Theta) = -\frac{1}{2}\mathbf{y}^\top \mathbf{K}_y^{-1}\mathbf{y} - \frac{1}{2}\log(|\mathbf{K}_y|) - \frac{N}{2}\log(2\pi) \quad (\text{C.3})$$

mit $\mathbf{K}_y = \mathbf{K}(\mathbf{U}, \mathbf{U}, \Theta) + \sigma_n^2 \mathbf{I}$.

Der Term \mathbf{K}_y bildet sowohl die verrauschten Messdaten \mathbf{y} als auch die rauschfreie Funktion $\mathbf{K}(\mathbf{U}, \mathbf{U}, \Theta)$ ab. Die Randverteilung in Gl. C.3 kann in drei verschiedene Unterterme aufgeteilt werden. Der erste Teil ist eine quadratische Funktion in \mathbf{y} , welche die Anpassung des Modells an die Trainingsdaten ausdrückt und somit einen komplexen Strafterm darstellt. Der zweite logarithmische, determinierte Term beschreibt die Komplexität des Modells abhängig von der Lage der Trainingsdaten im Faktorraum und des Hyperparametervektors Θ . Der dritte Term mit dem Multiplikator $\log(2\pi)/2$ ist ein konstanter, von der Messdatenanzahl N abhängiger Term zur Normierung [115, S. 300, 96, S. 29]. Die Hyperparameter werden mit dem Ziel die Randverteilung zu maximieren numerisch optimiert. Das Optimierungsproblem formuliert sich zu

$$\hat{\Theta}_{\text{ML}} = \arg \max_{\Theta} p(\mathbf{y}|\mathbf{U}, \Theta). \quad (\text{C.4})$$

Ein Maximum mit den optimierten Hyperparametern $\hat{\Theta}$ aus der Randverteilung beinhaltet automatisch einen Kompromiss zwischen der Modellgüte des ersten und der Komplexität des zweiten Terms in Gl. C.3. Jedes lokale Maximum erklärt eine eigene partikuläre Interpretation der Daten und die numerische Optimierung sollte aus Sicherheitsgründen mehrfach mit zufällig verteilten Startpunkten wiederholt und verglichen werden [115, S. 301, 96, S. 115]. Der Einfluss der Variation und die Einflussgüte der Hyperparameter auf die Modellvorhersage ist exemplarisch in [96, S. 20] visualisiert zu finden.

C.3 Hierarchische Optimierung von lokalen Modellnetzen

Die einzelnen Schritte des HILOMOT-Algorithmus sind nachfolgend von Hartmann [48, S. 76] übernommen sowie in Abb. C.2 aufgeführt:

1. *Initialisierung*: Startet mit einem einzigen LMN. Setze die Anzahl an LM auf $n_M = 1$. Die Gültigkeitsfunktion des im ersten Schritt globalen Modells LM ist über den kompletten Eingangsraum $\Phi_1 = 1$.
2. *Bestimme schlechtestes lokales Modell*: Berechnung der lokalen Verlustfunktion für alle $i = [1, \dots, n_M]$ lokalen Modelle mit der Verlustfunktion des gewichteten quadratischen Modellfehlers. Als Gewichtung dient der aktuelle Wert der Gültigkeitsfunktion Φ_i . Das LM mit dem höchsten Wert der Verlustfunktion wird mit dem Index l markiert.
3. *Teste verschiedene Teilungsrichtungen zur Initialisierung*: Das LM markiert mit l wird in zwei neue LM unterteilt. In jeder Dimension des Eingangsraums wird ein achsenorthogonaler Schnitt durch das LM-Zentrum durchgeführt. Die jeweilige globale Verlustfunktion wird für das gesamte Modell berechnet. Ist die Anzahl an LM $n_M > 1$, wird zusätzlich die Teilungsrichtung des Elternknotens getestet. Aus der besten Teilung folgt die Initialisierung der Optimierung.
4. *Nichtlineare Schnittoptimierung*: Vom Initialschnitt aus Schritt 3 beginnend werden Position und Richtung des neuen Schnitts bezüglich der Verlustfunktion des globalen Modells nichtlinear optimiert. Die Verlustfunktion ist die Wurzel des normierten quadratischen Fehlers (NRMSE). Nur der aktuellste Schnitt wird optimiert. Die restlichen Schnitte bleiben unverändert. Die Gesamtzahl der LM erhöht sich $n_M = n_M + 1$. Die Parameter der zwei neu erzeugten LM werden innerhalb jeder Berechnung der Verlustfunktion zur nichtlinearen Optimierung durch eine gewichtete LS Optimierung neu geschätzt.
5. *Überprüfe Abbruchkriterium*: Bei Erfüllung des Abbruchkriteriums findet keine weitere Teilung mehr statt und der Algorithmus ist beendet. Verwendet wird hier das korrigierte *Akaike information criterion* (AIC). Bei Nichterfüllung gehe zurück zu Schritt 2.

In Abb. C.2 sind die ersten drei Iterationen eines HILOMOT für einen zweidimensionalen Eingangsraum u_1 und u_2 schematisch nach [48, S. 76ff.] aufgezeigt. In der ersten Iteration startet der Algorithmus mit einem einzigen gültigen Modell 1–1. Für beide Eingangsdimensionen in u_1 - und u_2 -Richtung erfolgt der jeweils der achsenorthogonale hälftige Schnitt. Der Schnitt mit dem geringeren globalen Modellfehler wird als Initialisierung der nichtlinearen Schnittoptimierung ausgewählt. Anschließend an die Ausführung des Schnittes werden für beide LM 2–1 und 2–2 die lokalen Modellfehler berechnet. In Iteration zwei ist das schlechtere LM 2–2 grau

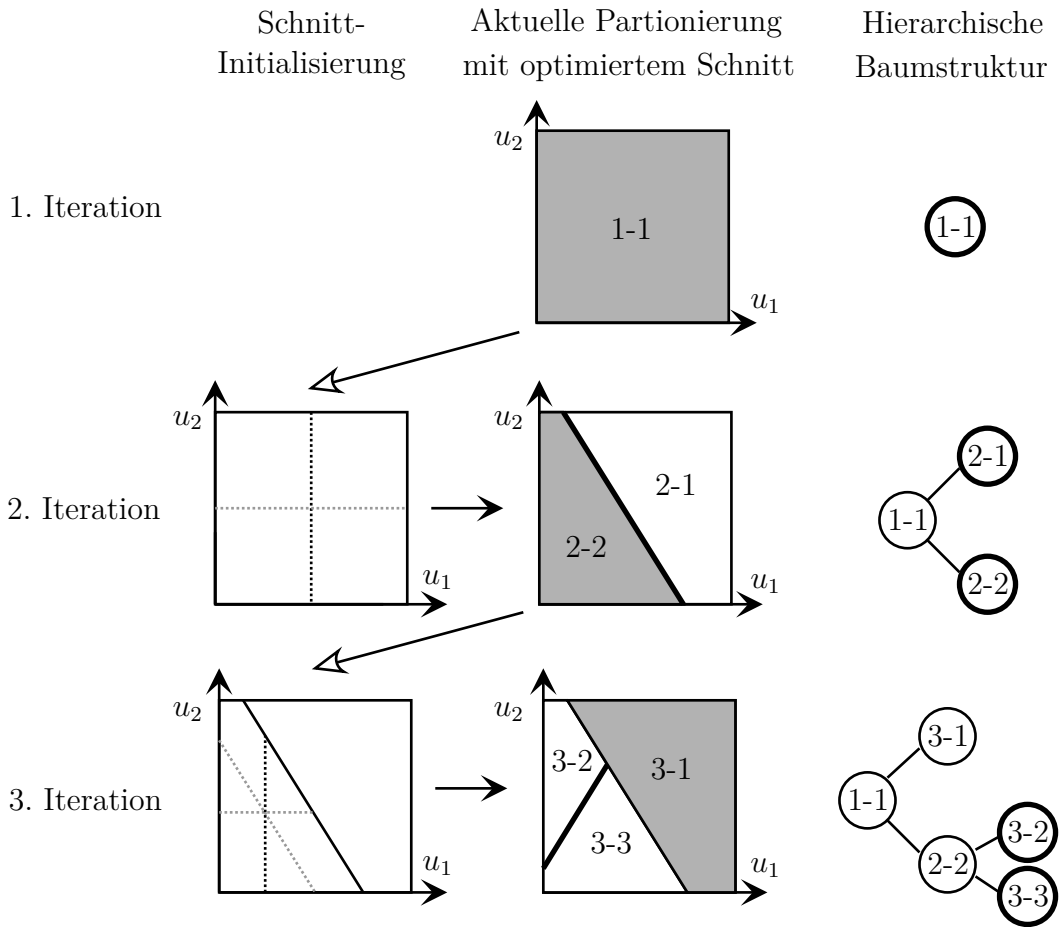


Abbildung C.2 Schnittinitialisierung, Partionierung und hierarchische Baumstruktur für drei mögliche Iterationen des HILOMOT Ansatzes. Das das schlechteste LM ist jeweils grau gekennzeichnet [48, S. 77]

markiert. Dieses gilt es im folgenden Schritt mit einem neuen LM weiter zu unterteilen, um den globalen Modellfehler weiter zu reduzieren. Der Algorithmus arbeitet iterativ und stoppt den Teilungsprozess erst, wenn das Abbruchkriterium erfüllt ist [48, S. 76].

C.4 Bestimmung von dynamischen Modellordnungen

Sollen dynamische Modellstrukturen parametrisiert werden, müssen zunächst die bedeutenden Regressoren selektiert werden, auf die das System sensitiv reagiert. Dies bietet eine besondere Herausforderung, da in den meisten Fällen kein Vorwissen über ein Prozessverhalten existiert. Eine Möglichkeit ist, Modelle mit immer weiter steigenden Modellordnungen zu identifizieren und anschließend die Modellfehler zu vergleichen. Ein Minimalwert einer Fehlerfunktion kann Aufschluss über die Modellordnung bringen. Die kann zu Überanpassungen führen. In diversen Studien wird für lineare Systeme eine steigende Modellkomplexität mit Straftermen versehen [66]. Die Regressorauswahl erfolgt bei linearen dynamische Modellstrukturen über Fehlerfunktionen in Abhängigkeit von der Modellordnung (siehe Abschnitt 3.3.2). Typischerweise besitzen beispielsweise die Modellausgänge $y(k-1)$ und $y(k-2)$ einen hohen statistischen Zusammenhang, müssen allerdings nicht zwingend relevant für den Prozess sein [85, S. 557]. Für eine steigende Anzahl an Regressoren wächst die Anzahl an zu schätzender Parametern an. Forschungsschwerpunkte liegen dabei nicht nur auf den Modellierungsansätzen, sondern auch in der Reduzierung der Parameter durch eine Vorauswahl der Regressoren.

Möglichkeiten zur Selektion entscheidender Regressoren (Modelleingänge) sind vorgestellt in [66, 67, 76]. Eingeführt wird die modellunabhängige Methode von [49]. Über die Relevanz von Eingängen entscheiden Lipschitz-Quotienten. Dieser Ansatz bietet den Vorteil die Systemordnung lediglich anhand von Messdaten abzuschätzen. Der Lipschitz-Quotient für den eindimensionalen Fall

$$l_{i,j} = \frac{|y(i) - y(j)|}{|\varphi(i) - \varphi(j)|}, \quad \text{für } i = [1, \dots, N], \quad j = [1, \dots, N] \quad \text{und} \quad (i \neq j). \quad (\text{C.5})$$

vergleicht die Abstände zwischen den Messungen y und den Regressoren φ eines jeden Datenpunktpaares i und j [85, S. 575]. Die Länge des Datensatzes beschreibt N und der Index i entspricht einer Position eines diskreten Signal. Der Lipschitz-Quotient bildet sich nach Gl. C.5 aus den jeweiligen Abständen zweier Punkte im Eingangs- $|y(i) - y(j)|$ sowie im Ausgangsraum $|\varphi(i) - \varphi(j)|$.

Die Idee der Lipschitz-Methode zur Bestimmung der Modellordnung leitet sich aus folgender Annahme ab. Ist der Abstand eines Signalpaares im Eingangsraum klein oder sogar identisch, dann kann davon ausgegangen werden, dass ein Punktepaar im Ausgang ebenfalls einen kleinen Abstand aufweist. Diese Theorie stützt sich auf der Annahme, dass sämtliche relevanten Eingänge zur Berechnung herangezogen sind. Für diesen Fall ergibt sich ein niedriger Lipschitz-Quotient. Fehlen jedoch entscheidende Eingänge in der Berechnung und die diskreten Werte des Ausgangspunktepaars liegen weit voneinander entfernt, ergibt sich ein hoher Zähler und der Lipschitz-Quotient steigt im Vergleich [49, 85, S. 575].

Da aus symmetrischen Gründen $l_{i,j} = l_{j,i}$ gilt, müssen $N(N - 1)/2$ Quotienten insgesamt berechnet werden [85, S. 576]. Der n_p -dimensionale Fall des Lipschitz-Quotienten

$$l_{i,j}^{(n_p)} = \frac{|y(i) - y(j)|}{\sqrt{(\varphi_1(i) - \varphi_1(j))^2 + \dots + (\varphi_{n_p}(i) - \varphi_{n_p}(j))^2}} \quad (\text{C.6})$$

beschreibt die Erweiterung der Gl. C.5 um einen Regressorvektor $\varphi(k)$ mit n_p relevanten Eingängen [85, S. 576]. Der Exponent n_p des Lipschitz-Quotient bezeichnet alle n_p relevanten Variablen, die aus dem Regressorvektor $\varphi(k)$ betrachtet sind. Iterativ wird bei der Lipschitz-Methode die Anzahl der Eingänge n_p erhöht und somit zusätzliche Eingänge in die Berechnung miteinbezogen. Die Systemordnung oder Dimension n_p wird dabei bis zur maximalen Anzahl an Systemeingängen erhöht und jeweils ein Lipschitz-Quotient berechnet. Für jede Dimension wird ein Lipschitz-Index berechnet und anschließend nach dem Wert sortiert und in einer Kurve aufgetragen.

Ein Lipschitz-Index L kann als maximal auftretender Lipschitz-Quotient der Dimension n_p definiert werden (siehe Gl. C.7) oder durch einen gewichteten, geometrischer Durchschnitt der k größten Lipschitz-Quotienten (Gl. C.8). Die k größten Lipschitz-Quotienten aller Einträge $l_{i,j}^{(n_p)}$ bezeichnet $l^{(n_p)}(k)$ für n_p Eingangsdimensionen [34, S. 104].

$$L^{(n_p)} = \max(l_{i,j}^{(n_p)}) \quad (\text{C.7})$$

$$L^{(n_p)} = \left(\prod_{k=1}^{n_q} \sqrt[n_p]{l^{(n_p)}(k)} \right)^{\frac{1}{n_q}} \quad (\text{C.8})$$

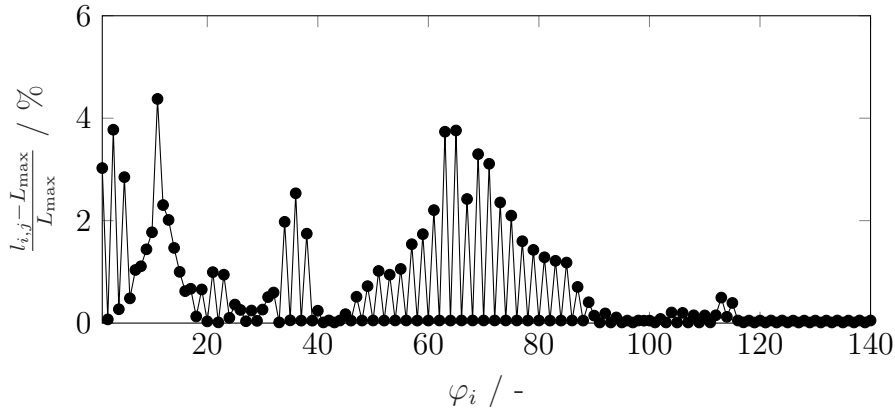


Abbildung C.3 Prozentuale Änderung des Lipschitz-Index für die Eingangsregressoren von der NFIR-Modellstruktur der nichtlinearen dynamischen Modellbildung für die Ausgangsgrößen \dot{m}_{NO_x} und \dot{m}_{HC}

Die Methode zur Bestimmung der Modellordnung mittels Lipschitz-Index ist modellunabhängig und ist besonders auf kleine Datensätze anwendbar. Mit steigender

Dimension der Eingangsgrößen steigt die Anzahl an Lipschitz-Quotienten kombinatorisch an. Die Lipschitz-Methode weist zwei weitere Nachteile auf. Erstere ist die Sensibilität gegenüber Prozessrauschen und der Datenverteilung. Die Rauschsensibilität kann mit einer Gewichtung $c > 1$ der Gl. C.8 eingeschränkt werden [85, S. 576, 13, S. 460]. Zweitens kann die Lipschitz-Methode nicht verwendet werden, um eine Totzeit zu identifizieren, da die Eingänge $u(k-1)$ automatisch mit einbezogen werden [85, S. 576, 66, S. 67, 67]. Allgemein gültig ist die Lipschitz-Methode zur Bestimmung von nicht-rekurrenten Modellstrukturen und deshalb geeignet für Modelle mit externer Ausgangsrückführung (NFIR und NARX), nicht aber für z. B. das NOE-Modell [85, S. 564].

Für das NFIR-Modell der der nichtlinearen dynamischen Modellbildung für die Ausgangsgrößen \dot{m}_{NO_x} und \dot{m}_{HC} aus Kap. 4.2.2 ist zur Bestimmung der Modellordnung der Lipschitz-Index herangezogen. Die prozentuale Änderung des Lipschitz-Index der NFIR-Eingangsregressoren φ_1 bis φ_{140} ist für das Fehlen des Regressoren φ_i in Abb. C.3 dargestellt. Zu Erkennen ist ein hoher Lipschitz-Index für die Regressoren $\varphi_i < 20$, für die Regressoren $\varphi_i < 40$ sowie für die Regressoren $40 < \varphi_i < 90$. Da bei $\varphi_i \approx 120$ ein lokales Minimum zu erkennen ist, wird die Anzahl an Eingängen bzw. die Anzahl an Regressoren auf $\varphi_i = 120$ festgelegt.

C.5 Numerische Optimierung

Liegt zur Abbildung des Systemverhaltens ein aus den Messdaten generiertes Modell vor, gilt es eine bestmögliche Einstellung der Eingangsparameter auf bestimmte Zielgrößen auszuwählen. Gegebenenfalls sind vorgegebenen Nebenbedingungen einzuhalten. Da die Anzahl an Parametern hoch und das Systemverhalten oftmals komplex und nichtlinear ist, werden für die Optimierung numerische Methoden eingesetzt. Eine Optimierung ermöglicht iterativ eine systematische Veränderung der Parameter im Eingangsraum einer Zielfunktion, um diese Zielfunktion zu minimieren [85, 59, 86, 115]. Oftmals sind Optimierungsprobleme klassifiziert als restringierte Minimierungsprobleme unter Berücksichtigung von (u. B. v.) Nebenbedingungen:

$$\arg \min_{\mathbf{u}} f(\mathbf{u}) \tag{C.9}$$

$$\text{u. B. v. } h(\mathbf{u}) = 0 \quad \text{und} \quad g(\mathbf{u}) \leq 0 \tag{C.10}$$

Die Nebenbedingungen werden als Gleichungs- h_i bzw. Ungleichungsnebenbedingung g_i bezeichnet. Ein Maximierungsproblem lässt sich durch Multiplikation mit (-1) in das äquivalente Minimierungsproblem transformieren. Ein numerischer Lösungsansatz ist je nach Eigenschaften der Zielfunktion auszuwählen. Auswahlkriterien sind z. B. Linearität, Stetigkeit, Unimodalität, stetige Differenzierbarkeit, Konvexität und Restriktion [59].

Ferner sind die numerischen Optimierungsansätze in lineare, nichtlinear lokale und nichtlinear globale Verfahren einteilbar [85, S. 25]. Lokale nichtlineare Optimierungs-

verfahren nutzen in der Regel die Differenzierbarkeit der Zielfunktion und wählen die Abstiegsrichtung um den betrachteten Punkt anhand dieser Informationen [59]. Diese gradientenbasierten Methoden iterieren zu einem Minimum aus einem Startpunkt heraus mit einem Gradientenabstieg. Existieren bei einer nicht-konvexen Gütefunktion viele lokale Minima, wird der Algorithmus mehrfach hintereinander gestartet. Mit zufällig gewählten Startpunkten findet ein lokales Optimierungsverfahren im Idealfall ein globales Minimum. Globale Optimierungsmethoden bieten den Vorteil, unabhängig der Problemstellung auf Kosten einer höheren Anzahl an Zielfunktionsauswertungen ein globales Optimum zu finden. Eine konvergente Lösung ist hierdurch im Allgemeinen in endlicher Zeit nicht garantiert [59]. Aus der Gruppe der globalen, nichtlinearen Optimierungsalgorithmen ist die Partikelschwarm- (PSO), die Innere-Punkte- und die multikriterielle Optimierung (GMO) in den Kap. C.5.1, C.5.3 und C.5.2 vorgestellt. Die Nomenklatur orientiert sich an der Literatur. Zudem wird für die Optimierung von mehreren Zielgrößen, die gleichzeitig zu optimieren sind, gezeigt, wie das Optimierungsproblem auf ein übergeordnetes abgebildet werden kann.

Sind bestimmte Faktorkombinationen durch die statistische Versuchsplanung nicht gemessen, muss der betrachtete Parameterraum verkleinert werden. Damit wird Extrapolation der Modellansätze außerhalb der Modellgültigkeit während der Optimierung verhindert. Daher wird der Versuchsraum der empirischen Modelle als Nebenbedingung in die Zielfunktion einbezogen. In der Regel werden konvexe Hüllen oder *Support vector machines* zur Bestimmung des erlaubten Parameterraum in der Optimierung verwendet [86, S. 86, 115, S. 456, 48, S. 106, 41].

C.5.1 Partikelschwarmoptimierung

Die Lösung eines lokalen bzw. deterministischen Verfahrens ist nicht in der Lage zwischen lokalem oder globalem Optimum zu unterscheiden. Diese Differenzierung gestaltet sich selten als trivial. Stochastische Lösungsansätze ermitteln globale Lösungen unabhängig von der Art des Optimierungsproblems. Diese globalen nichtlinearen Ansätze orientieren sich in der Regel an biologischen Naturphänomenen. Weitverbreitet sind hier die evolutionären Algorithmen, die der Evolution des Lebens nachempfunden sind und Schwarmintelligenzen, die das Schwarmverhalten von Tieren nachbilden. Für die modellbasierte Applikation wird die Partikelschwarm-Optimierung (PSO) als geeignetes, nichtlineares globales Verfahren vorgestellt. Die PSO bildet das Schwarmverhalten von Fischen, Insekten oder Vögeln simulativ nach [103, 57].

Bei der PSO erkunden eine Menge an Individuen eines Schwarms, bezeichnet als Partikel, den Parameterraum kommunikativ, ohne Information über die Gradienten auszutauschen. Nach der Auswertung der Zielfunktion vergleicht der Partikelschwarm die Information eines jeden Partikels über den individuellen Kostenwert und ermittelt den geringsten. Abhängig von den Ergebnissen der Zielfunktionsauswertungen

bewegt sich der ganze Schwarm in Richtung des Individuums mit dem geringsten Kostenwert. Mit dieser Metaheuristik wird der komplette Lösungsraum untersucht und ein globales Optimum gefunden [103]. Die Individuen werden zu Beginn zufällig im Parameterraum verteilt, wobei \mathbf{x}_k^i die aktuelle Position eines jeden Partikels i des Schwarms in der k -ten Iteration mit der zufälligen Geschwindigkeit \mathbf{v}_0 kennzeichnet. Die Geschwindigkeit stellt einen Betrag sowie eine Richtung dar. Nach den Auswertungen der Zielfunktionswerte kommunizieren die Individuen des Schwarms Position sowie Geschwindigkeit und passen abhängig von den Zielfunktionsauswertungen die Bewegung im Raum an. Die schrittweise Anpassung der Geschwindigkeit und der Position der Partikel ist mit den folgenden Iterationsgleichungen beschrieben [57]

$$\mathbf{x}_{k+1}^i = \mathbf{x}_k^i + \mathbf{v}_{k+1}^i, \quad (\text{C.11})$$

$$\mathbf{v}_{k+1}^i = \xi \mathbf{v}_k^i + h_1 r_1 (\mathbf{p}_k^i - \mathbf{x}_k^i) + h_2 r_2 (\mathbf{g}_k - \mathbf{x}_k^i). \quad (\text{C.12})$$

Die Partikelposition \mathbf{x}_{k+1}^i wird aus der aktuellen Position \mathbf{x}_k^i und dessen Geschwindigkeitsvektor \mathbf{v}_{k+1}^i berechnet. Die Bestimmung des Geschwindigkeitsvektors erfolgt über die folgenden drei Summanden: Die Trägheitskomponente mit der Trägheitskonstante ξ , der individuellen Erfahrung gewichtet mit h_1 und der gemeinsamen Erfahrung gewichtet mit h_2 . Der Ortsvektor des Bestwertes eines Partikel ist mit \mathbf{p}_k^i und der Ort des globalen Bestwertes eines Partikels mit \mathbf{g}_k gekennzeichnet. Die Konstanten r_1 und r_2 sind gleichverteilte Zufallszahlen der Menge $[0, 1]$ [111, S. 391]. Das Konvergenzverhalten und die Erkundung des Suchraums werden durch die Gewichtungsfaktoren ξ , h_1 und h_2 beeinflusst. Hohe Werte der Trägheitskonstante ξ begünstigen eine hohe Exploration des Suchraums, während die Konstanten h_1 und h_2 die Konvergenzgeschwindigkeit beeinflussen. Bei $h_1 = 0$ wird keine individuelle Erfahrung ausgetauscht. Es folgt eine schnelle Konvergenzgeschwindigkeit und ein erhöhtes Risiko, dass das Verfahren in lokalen Optima endet. Wird hingegen die Konstante $h_2 = 0$ gesetzt, existieren statt einem Partikelschwarm mit vielen Partikeln viele Schwärme mit nur einem Partikel [103]. Zusammenfassend wird ein jeder Partikel von einem Kompromiss aus individueller und globaler Erfahrung beeinflusst. Diese wird sich durch den Einfluss von Zufallszahlen und neuen Bestwerten iterativ anpassen. Über eine gewisse Rechenzeit ermitteln die Partikel optimale Bereiche und bewegen die Schwärme in diese Richtung [111, S. 391].

Eine geeignete Parametrierung und Implementierung der PSO für viele Applikationsaufgaben ist [103] entnommen. Der aus dieser Forschungsarbeit hervorgegangene *Matlab*-Algorithmus ist zur globalen Optimierung in dieser Arbeit verwendet. Bei dieser Variante ist der Schwerpunkt auf eine hohe Exploration des Parameterraums gelegt. Folglich ist der Wert der Trägheit der Partikel ξ hoch. Die Konstanten h_1 und h_2 sind niedrig gewählt, wobei der Fokus auf die individuelle Erfahrung mit $h_1 > h_2$ gelegt ist. Damit wird eine breite Verteilung der Parameter im Raum provoziert.

C.5.2 Multikriterielle Optimierung

Die PSO benötigt bei mehrdimensionalen Optimierungsproblemen eine Reduktion der Zielfunktion auf eine einzige übergeordnete Zielgröße. Die meisten Aufgaben der modellbasierten Applikation besitzen in der Regel mehrere gleichzeitig zu optimierende Zielfunktionen, die sich konträr verhalten können. Ein Beispiel ist der Kompromiss zwischen der Maximierung des Wirkungsgrades und der gleichzeitigen Emissionsminimierung eines Verbrennungsmotors. Ein multikriterielles Optimierungsproblem setzt sich zu einem Vektor aus mehreren zu optimierender Zielfunktionen zusammen [110, S. 170]

$$\arg \min f(x) = (f_1(x), \dots, f_p(x)). \quad (\text{C.13})$$

Sämtliche Optima definieren zusammen Punkte mit individuellen Lösungen, die zusammengefasst als Pareto-Grenze oder Pareto-Front bezeichnet werden. Eine Pareto-Front ist in Abb. C.4a für einen zweidimensionalen Lösungsraum dargestellt ist. Ein paretooptimales Ergebnis kennzeichnet ein individuelles Minimum auf einer Pareto-Grenze. Dieses kann nur auf Kosten einer Verschiebung einer anderen Zielgröße verringert werden. Eine optimale Lösung stellt einen akzeptablen Kompromiss zwischen den Zielgrößen dar und gibt eine bestimmte Faktoreinstellung wieder, die einem Punkt auf der Pareto-Front zugeordnet werden kann [115, S. 225, 110, S. 170].

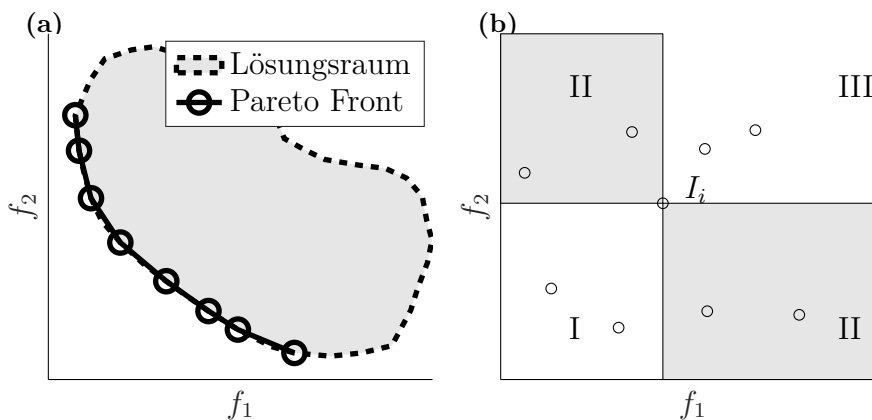


Abbildung C.4 Exemplarische Pareto-Front eines zweidimensionalen Lösungsraumes (a) und Dominanz eines Individuums I_i (b) nach [115, S. 227]

Paretooptimale Punkte sind definiert als Datenpunkte oder Individuen, die andere Datenpunkte dominieren und bezeichnet eine Teilmenge des Randes sämtlicher Zielfunktionslösungen. Ist ein Datenpunkt I_a mindestens in einer Zielgröße optimaler und in restlichen multikriteriellen Zielgrößen gleichwertig einem Punkt I_b , so dominiert Punkt I_a den Punkt I_b . Im Gegensatz dazu stellt Punkt I_b einen dominierten Punkt von I_a dar. In Abb. C.4b ist dies exemplarisch für den einzelnen Datenpunkt

I_i aufgezeigt. Sämtliche Punkte der Zielfunktionsauswertungen sind so eindeutig einem Bereich zuordenbar, je nach Aussage ob der Punkt dominiert (III), dominiert wird (I) oder keine Aussage darüber getroffen werden kann (II) [115, S. 327].

Genetische Optimierungsverfahren bieten bei komplexen Zusammenhängen zwischen Faktoren und Zielgrößen eine geeignete Methode zur Lösung multikriterieller Optimierungsprobleme und für die Ermittlung einer gleichmäßig verteilten Pareto-Grenze [115, S. 230]. Die der biologischen Evolutionstheorie nachempfundenen genetischen Algorithmen betrachten anstelle einer Lösung ganze Populationen von zulässigen Lösungen. Diese werden einzeln als Individuen unterschiedlicher Qualität beschrieben. Eine Population, die einen zeitlichen Aspekt berücksichtigt, wird als Generation angegeben. Das Ziel der Optimierung ist eine fortwährende Verbesserung der Qualität der Population mit den biologischen Operatoren: Kreuzung, Mutation und Selektion. Eine Gruppe von Individuen mit unterschiedlichen Faktoreinstellungen beschreibt eine zufälligen Startgeneration. Kreuzungen und Mischung der Erbinformationen selektiver Eltern-Individuen erzeugt Kinder. Mutation beeinflusst die Weiterentwicklung der Erbinformation eines Kindes. Dies hilft bei der Optimierung lokale Minima zu überwinden. Jedes Individuum wird mit einer Fitnessfunktion bewertet und die Selektion entscheidet, welche Individuen sich bei der Fortpflanzung durchsetzen. Selektion gibt damit die Suchrichtung des genetischen Optimierungsverfahren vor [110, S. 219, 111, S. 384, 115, S. 230]. Klassisch verwendet der genetische Algorithmus nach Goldberg [43] und Holland [53] eine binäre Kodierung. In dieser Arbeit wird allerdings der multikriterielle Optimierungsalgorithmus GAMULTIOBJ der *Matlab-Optimization-Toolbox* verwendet [122]. Dieser basiert auf reellwertigen Zahlen und dem Verfahren *Non-dominated sorting genetic algorithm* (NSGA-II) von Schoenauer u. a. [108, S. 182ff]. Der Schrittweise prinzipielle Ablaufplan des NSGA-II gliedert sich nach den folgenden Schritten [115, S. 236]:

1. Bildung einer Start-Generation G_0 aus n_g zufälligen Individuen.
2. Bestimmung des Ranges und der *Crowding distance* (CD) für jedes n_g Individuum.
3. Auswahl geeigneter Eltern E_i aus der aktuellen Generation mittels Rang und CD.
4. Berechnung der Kinder K_i durch Kreuzung der Eltern und mit anschließende Mutation.
5. Bestimmung des Ranges und der CD für alle Individuen ($G_i \cup K_i$).
6. Erzeugung einer neuen Generation G_{i+1} mittels Rang und CD aus $G_i \cup K_i$.
7. Wenn das Abbruchkriterium der Optimierung nicht erfüllt ist, gehe zurück zu Schritt 3.

Die Bestimmung des Ranges eines Individuums erfolgt nach einer Einteilung der Individuen in einzelne Fronten (siehe Abb. C.5). Die Pareto-Front beinhaltet sämt-

liche Individuen, die nicht von anderen dominiert werden, und erhält den Rang eins. Den Rang zwei bekommen die Punkte, welche eine neue Pareto-Front bilden, wenn sämtliche Punkte der Pareto-Front mit dem Rang eins nicht berücksichtigt werden [115, S. 236]. Weitere Ränge werden mit dem bilden neuer Pareto-Fronten ohne die bis dato in Ränge eingeteilten Individuen mit einem Sortieralgorithmus *Fast nondominated sorting approach* nach Schoenauer u. a. [108, S. 184] ermittelt.

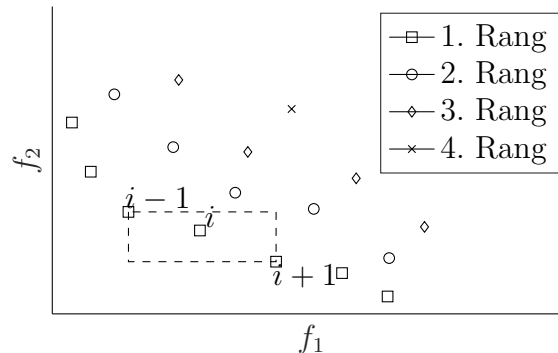


Abbildung C.5 Zweidimensionale Visualisierung der Einteilung von Individuen in den jeweiligen Rang eines zweidimensionalen Zielfunktionsraums und schematische Bestimmung der CD für ein Individuum I_i

Um eine breite Verteilung der einzelnen Optima des genetischen Optimierungsalgorithmus in der Pareto-Front zu erreichen, wird die CD für jedes Individuum als Distanzmaß bestimmt. Die CD definiert dabei ein Maß der Dichte von Lösungen der Individuen einer Generation. Im zweidimensionalen Funktionsraum von Abb. C.5 stellt die CD die Summe der Abstände zu benachbarten Individuen $i + 1$ und $i - 1$ dar. Dies entspricht den beiden Seitenlängen des dargestellten Rechtecks in Abb. C.5 [116, S. 82, 115, S. 237]. Beim NSGA-II werden geeignete Elternpaare aus zufälligen Individuen gewählt. Die Fitness besteht aus Rang und CD und wird anschließend bewertet. Treten zwei zufällige Individuen gegeneinander an, setzt sich das Individuum niedrigeren Ranges durch. Sind die Ränge zweier Individuen gleich, wird die Lösung mit größerer CD bevorzugt. Damit wird eine gleichmäßige Verteilung erzeugt. Randpunkte werden durch die Zuweisung CD „ist gleich unendlich“ immer vorgezogen. Sind genügend Elternpaare identifiziert, werden im nächsten Schritt durch Rekombination und Mutation neue Individuen erzeugt [115, S. 237, 108, S. 185, 116, S. 82].

C.5.3 Innere-Punkte Optimierung

Neben dem Verfahren der sequentiellen quadratischen Programmierung sind die Innere-Punkte Verfahren in der Lösung restringierter Minimierungsprobleme weit verbreitet [90, S. 117]. Die Idee von Innere-Punkte-Verfahren ist aus dem Inneren des zulässigen Bereichs das Minimum, welches auf dem Rand liegt, über einen zentralen

Pfad zu erreichen. Dazu wird eine Barrierefunktion mit einem logarithmischen Term auf die Ungleichungsnebenbedingungen eingesetzt, um diese in eine Gleichungsnebenbedingung zu transformieren. Die Ungleichungsnebenbedingung ist dadurch in jeder Iteration erfüllt und der Abstand der aktuellen Iteration zum Rand des zulässigen Bereiches wird von einem kleiner werdenden Barriereparameter kontrolliert.

Bei der Innere-Punkte Methode wird das nichtlineare Optimierungsproblem aus Gl. C.9 in der Ungleichungsnebenbedingung in Gl. C.10 durch eine Sequenz mit einer Barrierefunktion ersetzt, so dass die Form des nichtlinearen Optimierungsproblems als transformierten Form gegeben ist als [16, S. 1, 125, S. 393]

$$\arg \min_z \quad \varphi_\mu(\mathbf{z}) = f(\mathbf{x}) - \mu \sum_{i=1}^N \ln s_i, \quad \text{mit } \mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}, \quad (\text{C.14})$$

$$\text{u.B.v. } \mathbf{h}(\mathbf{x}) = \mathbf{0}, \quad \mathbf{g}(\mathbf{x}) + \mathbf{s} = \mathbf{0}. \quad (\text{C.15})$$

Hier ist $s \geq 0$ ein Vektor mit Schlupfvariablen und $\mu \geq 0$ der Barriereparameter. Für einen Wert von $\mu \rightarrow 0$ konvergiert die Lösung der Funktion $\varphi_\mu(\mathbf{z})$ gegen das Minimum des ursprünglichen, restringierten Problems. Wird die Randbedingung verletzt, strebt die logarithmische Barriere gegen Unendlich [90, S. 117]. Das Verfahren mit einer Barrierefunktion wird daher auch als inneres Penalty-Verfahren und die durch den Barriereparameter μ parametrisierte Trajektorie von Lösungen als *zentralen Pfad* bezeichnet [124, S. 4]. Die Lagrange Funktion auf das Optimierungsproblem angewendet ergibt

$$\mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}) = \varphi_\mu(\mathbf{z}) + \boldsymbol{\lambda}_h^\top \mathbf{h}(\mathbf{x}) + \boldsymbol{\lambda}_g^\top (\mathbf{g}(\mathbf{x}) + \mathbf{s}) \quad \text{mit } \boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}_h \\ \boldsymbol{\lambda}_g \end{bmatrix}, \quad (\text{C.16})$$

als Lagrange Multiplikatoren $\boldsymbol{\lambda} \in \mathbb{R}^N$ [16, S. 2, 125, S. 393]. Mit der Lagrange Funktion lassen sich die notwendigen Bedingungen der 1. Ordnung für das Barriereproblem mit dem Gradienten $\nabla_{\mathbf{z}, \boldsymbol{\lambda}} \mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}) = \mathbf{0}$ aufstellen. Dies ist auch als Karush-Kuhn-Tucker-Bedingungen (KKT) bekannt und gilt als notwendige Optimalitätsbedingung in der nichtlinearen Optimierung. Dieser lässt sich modifizieren indem die Spalte $\nabla_s \mathcal{L}$ zusätzlich mit der Diagonalmatrix S multipliziert wird. Das als modifizierte *primal-dual* KKT bezeichnete System von Gleichungen ist folglich beschrieben mit [16, S. 2, 125, S. 393, 124, S. 2, 107, S. 3]

$$\nabla \mathcal{L}(\mathbf{z}, \boldsymbol{\lambda}) = \mathbf{0} = \begin{bmatrix} \nabla_x f(\mathbf{x}) + \mathbf{J}_h(\mathbf{x})^\top \boldsymbol{\lambda}_h + \mathbf{J}_g(\mathbf{x})^\top \boldsymbol{\lambda}_g \\ \mathbf{S} \boldsymbol{\Lambda}_g - \mu \mathbf{e} \\ \mathbf{h}(\mathbf{x}) \\ \mathbf{g}(\mathbf{x}) + \mathbf{s} \end{bmatrix}. \quad (\text{C.17})$$

Hier sind $\mathbf{J}_h(\mathbf{x})$ und $\mathbf{J}_g(\mathbf{x})$ die Jacobi-Matrizen von $h(x)$ und respektive $g(x)$ differenziert nach x . S und $\boldsymbol{\Lambda}_g$ sind Diagonalmatrizen mit den Einträgen aus dem Schlupfvariablen-Vektor s bzw. dem Vektor der Lagrange Multiplikatoren $\boldsymbol{\lambda}_g$. Die Lösung des Gleichungssystems liegt mit der nicht-negativen Bedingung der Vektoren

s und λ immer strikt im Inneren des zulässigen Bereiches [90, S. 118]. Zur Lösung des Gleichungssystems C.17 wird das Newton-Verfahren zur Approximation verwendet [17, S. 155]. Das Gleichungssystem lässt sich formal aus einer Taylorreihenentwicklung von $\nabla \mathcal{L}$ im Punkt $[d_z^{(k)}, d_\lambda^{(k)}]$ umstellen, welches der Iterationsvorschrift des Newton Verfahrens entspricht, mit [69]

$$[\nabla^2 \mathcal{L}_k] \begin{bmatrix} d_z \\ d_\lambda \end{bmatrix} = -\nabla \mathcal{L}_k. \quad (\text{C.18})$$

Mit der Hesse-Matrix $\nabla_{xx}^2 \mathcal{L} = \nabla_{xx}^2 \mathcal{L}(\mathbf{z}, \boldsymbol{\lambda})$, den Gl. C.17 und deren Ableitung folgt das Gleichungssystem für das Newton-Verfahren zu [17, S. 155, 16, S. 6, 125, S. 393, 124, S. 3, 107, S. 3]

$$\begin{bmatrix} \nabla_{xx}^2 \mathcal{L} & 0 & \mathbf{J}_h(\mathbf{x}) & \mathbf{J}_g(\mathbf{x}) \\ 0 & \boldsymbol{\Lambda}_g & 0 & -\mathbf{S} \\ \mathbf{J}_h(\mathbf{x})^\top & 0 & 0 & 0 \\ \mathbf{J}_g(\mathbf{x})^\top & \mathbf{I} & 0 & 0 \end{bmatrix} \begin{bmatrix} d_x \\ d_s \\ d_{\lambda_h} \\ d_{\lambda_g} \end{bmatrix} = - \begin{bmatrix} \nabla_x f(\mathbf{x}) + \mathbf{J}_h(\mathbf{x})^\top \boldsymbol{\lambda}_h + \mathbf{J}_g(\mathbf{x})^\top \boldsymbol{\lambda}_g \\ \mathbf{S}_k \boldsymbol{\lambda}_g - \mu \mathbf{e} \\ \mathbf{h}(\mathbf{x}) \\ \mathbf{g}(\mathbf{x}) + \mathbf{s} \end{bmatrix}. \quad (\text{C.19})$$

Die Iterationsmatrix ist nicht-symmetrisch, kann aber durch Multiplikation der zweiten Reihe mit S^{-1} in eine symmetrische transformiert werden [124, S. 3]. In der Literatur ist für das symmetrische Gleichungssystem die Bezeichnung *primal-dual Iteration* verbreitet [107, S. 5]. Der aktuelle Wert der Iteration von λ_g steckt dabei nur in der Matrix $\boldsymbol{\Lambda}_g$ und der Hesse-Matrix $\nabla_{xx}^2 \mathcal{L}$ [18, S. 7]. \mathbf{I} bezeichnet hier die Einheitsmatrix. Der nächste resultierende Schritt der Iteration von den Optimierungsvariablen berechnet sich mit einer Schrittweite α durch [125, S. 393, 107, S. 3]

$$\mathbf{z}^+ = \mathbf{z} + \alpha_z d_z \quad \text{und} \quad \boldsymbol{\lambda}^+ = \boldsymbol{\lambda} + \alpha_\lambda d_\lambda. \quad (\text{C.20})$$

Dabei wird entlang der Newton Suchrichtung die Schrittweite so bestimmt, dass die Ungleichungen $s \geq 0$ und $\mu \geq 0$ immer erfüllt bleiben [90, S. 119]. Für die Berechnung der Schrittweiten α_z und α_λ werden zunächst die zwei maximalen Schrittweiten und anschließend die Schrittweite über ein *Backtracking*-Verfahren zu $\alpha_z \in (0, \alpha_z^{\max}]$ und $\alpha_\lambda \in (0, \alpha_\lambda^{\max}]$ bestimmt. Die Schrittweite wird in der Suchrichtung sukzessiv verkleinert, bis eine eindeutige Reduktion in einer sogenannten *Merit-Funktion* erreicht wird. Damit ist eine ausreichende Reduzierung in der Schrittweite vorgegeben. Die Iterationen der Gl. C.19-C.20 stellt die Basis für die Liniensuche des Innere-Punkte Verfahrens dar. Die Bestimmung einer effizienten Schrittweite erfolgt mit einem *trust-region-Ansatz*, bei dem die Barriere-Funktion aus Gl. C.14 um einen Strafparameter $\nu \geq 0$ erweitert wird [18, S. 8, 16, S. 3]. Die resultierende *Merit-Funktion* ist definiert als [125, S. 396, 16, S. 4]

$$\phi_\nu(\mathbf{z}) = \varphi_\mu(\mathbf{z}) + \nu \left\| \begin{bmatrix} \mathbf{h}(\mathbf{x}) \\ \mathbf{g}(\mathbf{x}) + \mathbf{s} \end{bmatrix} \right\|, \quad (\text{C.21})$$

mit der Euklidische Norm $\|\cdot\|$. Mit der Einführung der *Merit-Funktion* Gl. C.21 ist sichergestellt, dass der Newton Schritt in den Gleichungen C.17 konvergiert [124,

S. 4]. Für eine geeignete Wahl des Barriere-Parameter μ und einen iterativ angepassten Strafparameter ν existiert eine Sequenz an Schrittweiten α , so dass die Merit-Funktion $\phi_\nu(z)$ monoton in die vorgegebene Richtung d_z fällt [16, S. 4, 124, S. 4]. Die Schrittweite α_T wird dabei so gewählt, dass die *Armijo Bedingung*

$$\phi_\nu(\mathbf{z} + \alpha_T \alpha_z^{\max} \mathbf{d}_z) \leq \phi(\mathbf{z}) + \eta \alpha_T \alpha_\lambda^{\max} \mathbf{D}\phi_\nu(\mathbf{z}; \mathbf{d}_z) \quad (\text{C.22})$$

erfüllt ist [125, S. 398]. Hier stellt der Term $\mathbf{D}\phi_\nu(\mathbf{z}; \mathbf{d}_z)$ die Richtungsableitung der *Merit-Funktion* nach \mathbf{d}_z dar. Wenn ein nächster Schrittversuch des Newton-Verfahrens die *Merit-Funktion* nicht senkt, schließt der Algorithmus diesen aus und berechnet einen neuen Schritt [16, S. 4].

Der in jedem Schritt aktualisierte Barriere-Parameter μ beeinflusst die Konvergenz der Optimierung stark. Wird der Parameter zu schnell gesenkt, konvergieren die Variablen λ_g und s beide gegen Null und die Konvergenzgeschwindigkeit ist zu niedrig. Dies gilt auch bei zu langsamer Senkung des Parameters. Geeignete Abläufe und Algorithmen, welche die Barriere-Parameter und den Strafparameter aktualisieren und anpassen, sind beschrieben in [125, S. 399]. Ein Ablaufplan bzw. Pseudocode des Innere Punkte Verfahrens ist ebenfalls detailliert zu finden in [125, S. 395]. In *Matlab* ist das Innere-Punkte Verfahren dem Anwender in der *Optimization Toolbox* als *fmincon* Lösungsalgorithmus zur Verfügung gestellt und findet in dieser Arbeit Verwendung [122].

C.5.4 Formulierung der Kostenfunktion

Werden bei der Optimierung mehrdimensionale Zielgrößen definiert, die gleichzeitig zu optimiert sind, sind die Zielgrößen auf eine einzige, übergeordnete zu reduzieren. Unterscheiden sich einheitenbehaftete Eingangs- oder Ausgangsgrößen in der Größenordnung, werden die Größen zwischen $[-1, 1]$ skaliert oder normiert $[0, 1]$ und damit vom Einfluss einer Einheit befreit [69, S. 61]. Das verbessert das Konvergenzverhalten. Die Optimierung mehrere verschiedener Zielfunktionen kann mit einer übergeordneten Zielgröße z_{ZF}

$$z_{ZF} = \mathbf{w} \cdot \mathbf{z} = w_1 z_1 + w_2 z_2 + \dots + w_K z_K, \quad \text{mit} \quad \sum_{i=1}^K w_i = 1 \quad (\text{C.23})$$

als eine mit \mathbf{w} gewichteten Summe der zu optimierenden Parameter \mathbf{z} erfolgen [115, S. 333]. Die Anzahl an Zielgrößen stellt K dar. Die Zielfunktion hängt direkt vom Verhältnis der einzelnen Gewichtungen ab. Eine geeignete Kombination der Gewichte kann grundsätzlich jeden Punkt einer konvexen Pareto-Front ermitteln. Eine sinnvolle Gewichtung der einzelnen Zielgrößen muss vor der Optimierung getroffen werden. Eine Zielgröße mit hohem Einfluss auf die übergeordneten Zielfunktion kann klein gewichtet von einer unwesentlichen Zielgröße verschleiert werden, die zu hoch gewichtet ist. Wird von einer Ursprungsapplikation ausgegangen, kann eine Optimierung zur Verbesserung der Basisapplikation in den Ungleichungsnebenbedingungen

definiert werden. Eine Zielgröße wird unter den Restriktionen der restlichen Zielgrößen optimiert. Zur Verbindung mehrerer Zielfunktionen zu einer einzigen skalaren Funktion können Lagrange-Multiplikatoren herangezogen werden [84, S. 25, 69, S. 62].

C.6 Fehlerfunktion des Gütekriteriums der Kennfelderstellung

Die Fehlerfunktion e_p bildet sich aus der Differenz der N_{BP} vorgegebenen, tatsächlichen Werten der Betriebspunkte \mathbf{v} und den interpolierten Werten $\hat{\mathbf{v}}_p$. Eine Minimierung der quadrierten Differenz erfolgt über die Parameterschätzung mittels der kleinste Quadrate Methode (LS) zu dem Gütekriterium c_{LS} .

$$\arg \min_{\mathbf{z}} c_{\text{LS}}(\mathbf{z}) = \mathbf{e}_p^\top \mathbf{e}_p \quad \text{mit} \quad \mathbf{e}_p = \mathbf{v} - \hat{\mathbf{v}}_p = \mathbf{v} - \mathbf{M}\mathbf{z}. \quad (\text{C.24})$$

Die Interpolationsmatrix hat die Dimension $N_{\text{BP}} \times N_{\text{KP}}$. Mit dieser Formulierung lassen sich gesuchten Kennfeldstützstellen aus den Betriebspunkten direkt analytisch berechnen. Hierfür werden die Beziehungen multipliziert, die einzelnen Terme abgeleitet und anschließend gleich Null gesetzt. Eine Auflösung nach $\hat{\mathbf{z}}$ führt auf die Schätzgleichung

$$\hat{\mathbf{z}} = (\mathbf{M}^\top \mathbf{M})^{-1} \mathbf{M}^\top \mathbf{v}. \quad (\text{C.25})$$

Die Schätzgleichung C.25 ist nur für den Fall $\det(\mathbf{M}^\top \mathbf{M}) \neq 0$ lösbar, wenn die Inverse zu \mathbf{M} existiert und der Term $(\mathbf{M}^\top \mathbf{M})$ nicht singular ist. Für den Rang muss $\text{Rang}(\mathbf{M}) \leq \min(N_{\text{BP}}, N_{\text{KP}})$ gelten. Eine einfach zu bestimmende exakte Lösung gibt es dabei für den Fall, dass die Anzahl an Betriebspunkten gleich der Anzahl zu bestimmender Kennlinien- oder Kennfeldstützstellen ist $N_{\text{BP}} = N_{\text{KP}}$. Sind mehr Betriebspunkte als Stützstellen vorhanden, gilt es eine Lösung der optimalen Kennlinie zu approximieren mittels Gl. C.25. Im Extremfall, bei dem ein Kennfeldstützstellenwert mehrdeutig bestimmt werden kann, bietet die Gl. C.25 eine nicht eindeutigen Lösung, da der Rang der Matrix kleiner ist als die Anzahl ihrer Spalten und damit kleiner als die Anzahl zu bestimmender Kennfeldstützstellenwerte. Dieser unterbestimmte Fall als zusätzlicher Freiheitsgrad zieht eine beliebige Lösung bei der Berechnung eines Kennfeldes nach sich, welcher mit der Einführung eines Kriteriums für die Kennfeldglattheit als Verfahren für eine eindeutige Lösung erweitert wird. Dies ist in Abb. 2.8a dargestellt. In Abb. 2.8 sind Beispiellösungen für eine Kennlinienerstellung an den jeweiligen Stützstellen für bestimmte Betriebspunkte gegeben.

Mit der Einführung der Kennfeldglattheit als Gütemaß sind die einzelnen Terme der Kennfelderstellung vollständig beschrieben. Eine gewichtete Lösung für ein komplett Glatte Kennfeld ergibt sich für den Fall $\zeta = 0$ und die Lösung entspricht einer einfachen, planaren Ebene ohne jegliche Steigung und Krümmung. Die exakte Lösung

Tabelle C.1 Lösungsszenario für den Kennfeldvektor \hat{z} [30, S. 11]

Fall	Lösung
$\text{Rang}(\mathbf{M}) = N_{\text{KP}} \ \& \ N_{\text{BP}} = N_{\text{KP}}$	exakt
$\text{Rang}(\mathbf{M}) = N_{\text{KP}} \ \& \ N_{\text{BP}} > N_{\text{KP}}$	nicht exakt
$\text{Rang}(\mathbf{M}) < N_{\text{KP}} \ \& \ N_{\text{BP}} < N_{\text{KP}}$	nicht eindeutig lösbar

$\zeta = 1$ kann ein raues Kennfeld hoher Gradienten erzeugen. Mit der Einführung der analytischen Kennfeldglattheit ist die Anzahl zu optimierender Parameter in der Kennfeldoptimierung deutlich reduziert [69, S. 71]. Werden die beiden Lösungen Gl. C.25 für die Exaktheit und Gl. 2.29 für die Glattheit in diese Kostenfunktion eingesetzt, ergibt sich das vollständige Minimierungsproblem zu

$$\arg \min_z c_{\text{KF}}(\mathbf{z}) = \zeta(\mathbf{v} - \mathbf{M}\mathbf{z})^\top(\mathbf{v} - \mathbf{M}\mathbf{z}) + (1 - \zeta)\mathbf{z}^\top \mathbf{R}_d \mathbf{z}. \quad (\text{C.26})$$

Für die Schätzgleichung der Kennfeldstützstellen wird der Ausdruck in Gl. 2.29 nach \mathbf{z} abgeleitet und gleich Null gesetzt. Durch Umformung ergeben sich

$$\underbrace{(\zeta \mathbf{M}^\top \mathbf{M} + (1 - \zeta) \mathbf{R}_d)}_{\mathbf{H}} \hat{\mathbf{z}} = \underbrace{\zeta \mathbf{M}^\top \mathbf{v}}_{\mathbf{f}}, \quad (\text{C.27})$$

$$\hat{\mathbf{z}} = \mathbf{H}^{-1} \mathbf{f}(\mathbf{v}). \quad (\text{C.28})$$

C.7 Systemidentifikation linearer dynamischer Modelle

Für die Totzeitschätzung mittels Verfahren der expliziter Parameterschätzung ist dem interessierten Leser eine kurze Einführung in die Systemidentifikation linearer dynamischer Modelle gegeben. Sind in Kap. 2.2 nichtlineare dynamische Ansätze zur Abbildung des Systemverhaltens vorgestellt, werden zur Schätzung von Totzeitgliedern werden hier deren lineare Ansätze beschrieben. Die hier eingeführten Modellbeschreibungen stellen Differenzgleichungen dar und der Begriff Differenzgleichungsmodell beschreibt das Verhalten des Systems von Eingang zu Ausgang in Abhängigkeit von zurückliegenden Werten der Ausgangs- und der Eingangsgröße und gemessener oder beobachteter Störgröße [12, S. 292, 72, S. 1]. Zur Herleitung der in dieser Arbeit verwendeten linearen Modellstrukturen der Systemidentifikation dient Abb. C.6. Hier ist der allgemeine Modellansatz für einen gestörten linearen Prozess dargestellt, bei dem der Prozess von einem ungestörten Anteil, einem reinen Störterm, der das Prozessrauschen beschreibt und einer Eigendynamik beschrieben ist. Das System wird von der Eingangsgröße $u(k)$ angeregt und gibt mit einem externen Störsignal e eine Ausgangsgröße $y(k)$ aus. Das Argument $u(k)$ steht in diesem Fall

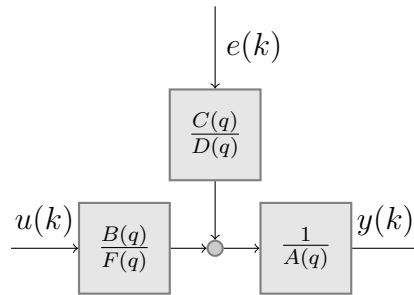


Abbildung C.6 Allgemeine Darstellung einer vollständigen linear dynamischen Modellstruktur für einen gestörten linearen Prozess nach [85, S. 467]

vielmehr für den gesamten Verlauf der Eingangsgröße bzw. für die Eingangsfunktion [12]. Der rauschfreie Term lässt sich mit der Gl. C.29 diskreten Übertragungsfunktion mit dem Verschiebeoperator q der vereinfachten Transformation $q^{-i}u(k) = u(k - i)$ darstellen [85, S. 464]. Die Transformation wandelt die Differenzgleichung in eine diskrete Übertragungsfunktion und der Multiplikation mit z^{-i} entspricht einer Verzögerung von i Zeitschritten. Die Konstanten n_b und n_f stellen dabei die Ordnungen des Zähler- bzw. Nennerpolynoms dar, n_k die Anzahl an Totzeitgliedern am Eingang. Analog kann der Stör- oder Rauschterm in Gl. C.30 angegeben werden.

$$G(q) = \frac{B(q)}{F(q)} = \frac{b_1 z^{-n_k} + b_2 z^{-n_k-1} + \dots + b_{n_b} z^{-n_k-n_b+1}}{1 + f_1 z^{-1} + \dots + f_{n_f} z^{-n_f}} \quad (\text{C.29})$$

$$H(q) = \frac{C(q)}{D(q)} = \frac{1 + c_1 z^{-1} + \dots + c_{n_c} z^{-n_c}}{1 + d_1 z^{-1} + \dots + d_{n_d} z^{-n_d}} \quad (\text{C.30})$$

Das vollständige Prozessmodell kann durch Kombination einer deterministischen $G(q)$ und einer stochastischen Transferfunktion komplett modelliert werden. Dadurch ist der Modellausgang eines linearen System $y(k)$ zum Zeitpunkt k durch eine lineare Filterung $G(q)$ des Eingangs $u(k)$ und einer lineare Filterung $H(q)$ des Rauschterms $e(k)$ unter Berücksichtigung eines weiteren Nennerpolynoms $A(q)$ vollkommen beschrieben. Die vollständige Prozessbeschreibung erfolgt mit dem Polynom $A(q)$ zur Beschreibung der Eigendynamik zu

$$A(q)Y = G(q)U + H(q)E. \quad (\text{C.31})$$

Die gesamte Terminologie der dynamischen Modellstrukturen folgt dabei standardisiert dem Werk von [73]. Für bestimmte Annahmen die Polynome A , B , C , D und F lassen sich mit diesen Transferfunktionen sämtliche bekannte linear dynamischen Modellstrukturen abbilden. Zur Herleitungen, Terminologie und Klassifizierung der verschiedenen Modellstrukturen oder Zeitreihenmodelle sind zum Einstieg die Werke [73, 85, 12, 72] zu empfehlen. Ziel der Systemidentifikation ist es eine passende Modellstruktur zu definieren und die Parameter des Modells zu schätzen. Je nach Modellwahl enthält ein entsprechender Parametervektor Θ die Koeffizienten a_i , b_i , c_i , d_i und f_i aus den Übertragungsfunktionen, welche aus den Messdaten approximiert werden müssen.

C.8 Modellgleichungen der Zufallssignale der Totzeitschätzung

Vorgestellt sind in diesem Abschnitt die beiden Zustandsraummodellgleichungen der linearen und nichtlinearen Signalgenerierung des Beispiels aus 3.3.3.

Das lineare Zustandsraummodell

$$\begin{aligned} \begin{bmatrix} \dot{x}_1(k) \\ \dot{x}_2(k) \end{bmatrix} &= \begin{bmatrix} -14,27 & -26,87 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(k) \\ y_L(k) &= \begin{bmatrix} 4,03 & 45,10 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + u(k) \end{aligned} \tag{C.32}$$

transformiert das Zufallssignal $u(k)$ auf das lineare Ausgangssignal (vgl. Gl. 3.27). Ein lineares Zustandsraummodell nach Gl. 3.27 kann zu einem nichtlinearen Gl. C.33 erweitert werden [85]. Das nichtlineare Ausgangssignal ergibt sich aus der Transformation des Zufallssignals $u(k)$ mit Gl. C.34.

$$\begin{aligned} \dot{\mathbf{x}}(k) &= \mathbf{h}(\mathbf{x}(k), u(k)) \\ y(k) &= g(\mathbf{x}(k)) \end{aligned} \tag{C.33}$$

$$\begin{aligned} \begin{bmatrix} \dot{x}_1(k) \\ \dot{x}_2(k) \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ -0,1 & -0,075 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k)^2 \\ y_{NL}(k) &= 2 \cos(x_1(k)) + \frac{1}{3} \sin(x_2(k)) \end{aligned} \tag{C.34}$$

Wissenschaftliche Publikationen

- 2020** Methods of Sensitivity Analysis in model-based Calibration, Automotive and Engine Technology
- 2019** Vortrag Methods of Sensitivity Analysis, WKM Symposium, Karlsruhe
- 2019** Time-delay-estimation for automotive applications, International Conference on Calibration, Methods and Automotive Data Analytics
- 2019** Vortrag Time-delay-estimation for automotive applications, International IAV Conference, Berlin