

# Emission Modelling and Model-Based Optimisation of the Engine Control

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

von

**Dipl.-Math. Heiko Sequenz**

geboren am 29. April 1983 in Groß-Gerau

Referent: Prof. Dr.-Ing. Dr. h. c. Rolf Isermann  
Korreferenten: Prof. Dr. rer. nat. Michael Kohler  
Prof. Dr.-Ing. Marius Pesavento

Tag der Einreichung: 22. Oktober 2012  
Tag der mündlichen Prüfung: 25. Februar 2013



D 17

Darmstädter Dissertationen



---

# Danksagung

Zunächst danke ich Prof. Isermann der durch die Finanzierung und durch zahlreiche Diskussionen zu meiner Forschungstätigkeit die Erstellung dieser Dissertation ermöglichte. Dank ihm erhielt ich einen tiefen Einblick in Forschungsthemen der Automobilindustrie, welche mich für meinen weiteren beruflichen Weg geprägt haben. Insbesondere danke ich ihm auch für die Möglichkeiten meine wissenschaftliche Ergebnisse auf nationalen und internationalen Konferenzen präsentieren zu können.

Des Weiteren danke ich Prof. Kohler und Prof. Pesavento für die Übernahme des Korreferats. Ich freue mich, dass mit Prof. Kohler ein Korreferent des Fachbereichs Mathematik meine Arbeit von der mathematischen Seite bewertet hat, da dies die Grundlage dieser Arbeit und auch mein persönlicher Hintergrund ist. Auch Prof. Pesavento danke ich für die gewinnbringenden Anmerkungen zu den verwandten Modellbildungsverfahren.

Ich möchte auch den Kollegen vom Institut für die stets angenehme Arbeitsatmosphäre danken, da gute Arbeit nur in einer solchen entstehen kann. Bei diesem Dank schließe ich sowohl die wissenschaftlichen Mitarbeiter als auch die administrativen Angestellten mit ein, welche stets mit Rat und Tat im täglichen Arbeitsablauf zur Seite standen. Ich möchte an dieser Stelle insbesondere Matthias Mrosek danken, mit dem ich zahlreiche Diskussionen zu meiner Arbeit hatte und mit dem ich auch einige gemeinsame Veröffentlichungen erstellen konnte. Auch danke ich Ralf Zimmerschied, Sebastian Clever, Karl von Pfeil und Martin Kohlhase die mir insbesondere in meiner Anfangszeit bei der Orientierung für meine Arbeit geholfen haben. Marek Kowalczyk, Markus Bauer, Christopher Eck, Andreas Sidorow, Sebastian Zahn, Mark Beck und Florian Kunkel danke ich ebenso für die vielen Diskussionen und die gute Zeit am Institut.

An dieser Stelle will ich auch insbesondere meinen Eltern von ganzem Herzen danken, da sie aufgrund zahlloser Gründe die Basis für diese Arbeit sind. Sie unterstützten mich in vielerlei Hinsicht auf meinem Weg zu dieser Dissertation und haben stets ein offenes Ohr. Auch meiner Frau Nadine bin ich für ihre liebevolle Unterstützung und das Verständnis beim Schreiben meiner Dissertation zutiefst dankbar. Sie ist ein bedingungsloser Rückhalt und hat mich nicht nur in dieser Zeit stetig unterstützt. Auch meinem Bruder Oliver möchte ich danken, der mich als Bruder und Freund auf dem Weg zu dieser Arbeit geprägt hat.

*Für meine Eltern  
und für Nadine*

# Contents

<b>Symbols and Abbreviations</b>	<b>VIII</b>
<b>Abstract</b>	<b>XI</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Model-based optimisation of the engine control . . . . .	1
1.2 Previous work . . . . .	4
1.3 Thesis outline and new contributions . . . . .	6
<b>2 Model Structures</b>	<b>8</b>
2.1 Look-up tables . . . . .	9
2.1.1 Identification . . . . .	11
2.1.2 Regularisation . . . . .	13
2.1.3 Combination of two-dimensional look-up tables . . . . .	16
2.1.4 Inversion of look-up tables . . . . .	20
2.2 Polynomial models and Design of Experiments . . . . .	22
2.2.1 Selection of regressors . . . . .	24
2.2.2 Design of Experiments . . . . .	26
2.3 LOLIMOT . . . . .	29
2.4 LOPOMOT . . . . .	31
2.4.1 Local selection of regressors . . . . .	32
2.4.2 Global selection of model partition . . . . .	34
2.4.3 Dynamic models . . . . .	34
2.4.4 Comparison to LOLIMOT . . . . .	36
2.5 Kernel models . . . . .	40
2.6 Summary . . . . .	41
<b>3 Static and Dynamic Combustion Models</b>	<b>43</b>
3.1 Model inputs and outputs . . . . .	47
3.1.1 Model inputs . . . . .	47
3.1.2 Model outputs . . . . .	51
3.2 Local and global model structures . . . . .	53
3.3 Measurement design . . . . .	56
3.4 Training and validation results . . . . .	59
3.4.1 Results from stationary data . . . . .	59
3.4.2 Dynamic model results . . . . .	61
3.5 Analysis of the stationary emission characteristics . . . . .	64
3.5.1 Intersection plot tool . . . . .	64

3.5.2	NO <sub>x</sub> -soot trade-off . . . . .	66
3.6	Analysis of the dynamic emission characteristics . . . . .	67
3.6.1	Step in injection quantity and ramp in engine speed . . . . .	67
3.6.2	Influences of measurement dynamics . . . . .	69
3.7	Summary . . . . .	72
<b>4</b>	<b>Optimisation of the Stationary Engine Operation</b>	<b>74</b>
4.1	Objective function . . . . .	75
4.1.1	Criterion for a local optimisation . . . . .	75
4.1.2	Criterion for a global optimisation . . . . .	78
4.2	Local optimisation . . . . .	80
4.2.1	Numerical optimisation algorithms . . . . .	80
4.2.2	Local optimisation results . . . . .	81
4.3	Global optimisation . . . . .	85
4.3.1	Robust optimisation . . . . .	86
4.3.2	Smoothing of engine maps . . . . .	87
4.4	Discussion of optimisation results . . . . .	89
4.4.1	Comparison of the local and global optimisation results . . . . .	90
4.4.2	Simulation of an engine series production . . . . .	91
4.4.3	Dynamic simulation of the NEDC . . . . .	93
4.5	Summary . . . . .	95
<b>5</b>	<b>Optimisation of the Dynamic Engine Operation</b>	<b>97</b>
5.1	Objective function . . . . .	98
5.2	Offline optimisation of an acceleration event . . . . .	99
5.3	Closed loop controls . . . . .	102
5.4	Model-based smoke limitation . . . . .	106
5.4.1	Functional structure of the smoke limitation . . . . .	108
5.4.2	Implementation in the ECU . . . . .	109
5.4.3	Test bed results . . . . .	111
5.5	Summary . . . . .	114
<b>6</b>	<b>Comparison of Model Structures for Control Applications in the ECU</b>	<b>116</b>
6.1	Comparison of model structures for a NO <sub>x</sub> model . . . . .	116
6.1.1	Model accuracy . . . . .	117
6.1.2	Computing time . . . . .	121
6.1.3	Memory demand . . . . .	124
6.1.4	Conclusions . . . . .	125
6.2	Implementation of a torque model in the ECU . . . . .	127
6.2.1	Look-up table structure for the torque model . . . . .	128
6.2.2	Injective mapping constraint for the torque model . . . . .	130
6.2.3	Inversion of the torque model . . . . .	132

---

6.3	Open loop control of the smoke limitation . . . . .	133
6.4	Summary . . . . .	136
<b>7</b>	<b>Conclusions</b>	<b>137</b>
<b>A</b>	<b>Air Path Model</b>	<b>141</b>
A.1	Air path states . . . . .	142
A.2	Gas composition dynamics . . . . .	144
<b>B</b>	<b>Combustion Models Appendix</b>	<b>146</b>
B.1	Combustion model database . . . . .	146
B.2	Models without in-cylinder pressure measurements . . . . .	146
B.3	Computing the global-local model structure . . . . .	147
B.4	Dynamic emission characteristics . . . . .	148
<b>C</b>	<b>Engine Test Bed</b>	<b>152</b>
<b>D</b>	<b>Mathematical Appendix</b>	<b>155</b>
D.1	Recursive least squares algorithm . . . . .	155
D.2	Regularisation for look-up tables . . . . .	156
D.3	Pyramidal weighting function . . . . .	158
D.4	LOPOMOT c-function . . . . .	158
D.5	Derivation of the weights for the multi-objective optimisation function . . . . .	158
	<b>Bibliography</b>	<b>160</b>

# Symbols and Abbreviations

## Symbols

$\mathcal{A}$	best set of potential regressors
$\mathcal{B}$	best set of measurements
bsfc	brake specific fuel consumption in g/kWh
bsfc <sub>inner</sub>	inner brake specific fuel consumption in g/kWh
$c_{\text{nox}}$	concentration of NO <sub>x</sub> in ppm
$c_{\text{opa}}$	concentration of opacity in %
$C_p$	criterion of fit value (Mallows $C_p$ statistic)
$c_{\text{soot}}$	concentration of soot in mg/m <sup>3</sup>
$\mathcal{D}$	drivability space
$\Delta m_{\text{dyn}}$	additional dynamic emissions in mg/km
$\Delta \dot{m}_{\text{mod}}$	variations due to model uncertainties in mg/s
$\Delta \dot{m}_{\text{series}}$	variations due to series variations in mg/s
$e, \mathbf{e}$	error (vector)
$G$	transfer function
$\mathbf{\Gamma}$	regularisation matrix
$\gamma$	adaption weight for the recursive least squares algorithm
$\mathbf{H}$	hat matrix of a linear model, such that $\mathbf{H}\mathbf{y} = \hat{\mathbf{y}}$
$h, \mathbf{h}$	bandwidth (vector)
$i, j$	counting variables
$J$	loss function
$K_h$	Kernel function
$k$	discrete time / number of potential regressors
$k_j$	weights for the engine operation points
$k_{\text{nox}}, k_{\text{soot}}$	weights of an objective function for the emissions
LHV <sub>Diesel</sub>	lower heating value of Diesel
L <sub>st</sub>	stoichiometric air to fuel ratio (14.545)
$\lambda$	regularisation parameter / air to fuel ratio
$M$	number of local models
$M_{\text{eng}}$	engine torque in Nm
$\dot{m}_{\text{air}}$	air mass flow rate in kg/h
$m_{\text{air}}$	air mass per cycle in mg/cyc
$m_{\text{egr}}, m_{\text{lp-egr}}$	(low-pressure) egr mass per cycle in mg/cyc
$\dot{m}_{\text{egr}}, \dot{m}_{\text{lp-egr}}$	(low-pressure) egr mass flow rate in kg/h
$m_{\text{eng,in}}$	total gas mass per cycle in mg/cyc
$\dot{m}_{\text{f}}$	fuel mass flow rate in mg/s
$\dot{m}_{\text{nox}}$	emission mass flow rate of NO <sub>x</sub> in mg/s

$\dot{m}_{\text{soot}}$	emission mass flow rate of soot in mg/s
$m_u, m_y$	dynamical order of the input, output
$\mathcal{M}$	set of potential regressors
$N$	number of measurements
$\mathcal{N}$	set of potential measurements / normal distribution
$\text{NO}_{x,\text{limit}}$	emission standard limit for $\text{NO}_x$ in mg/km
$n$	number of selected regressors
$n_{\text{eff}}$	effective number of selected regressors
$n_{\text{eng}}$	engine speed in rpm
$o$	order of a polynomial
$\omega_{\text{eng}}$	engine speed in radian
$p_{2i}$	intake pressure in bar
$p_{\text{rail}}$	rail pressure in bar
$\Phi$	weighting function
$\varphi_{\text{mi}}$	crank angle of main injection in °CA
$\varphi_{\text{pi}}$	crank angle of pilot injection in °CA
$\varphi_{\text{Q50}}$	crank angle of 50 % mass fraction burnt in °CA
<b>Q</b>	weighting matrix
$q_{\text{pi}}$	quantity of pilot injection in mg/cyc
$R^2$	coefficient of determination
<b>RMSE</b>	root mean square error
$R_o$	remainder of an approximation
$\rho_{\text{Diesel}}$	density of Diesel in mg/mm <sup>3</sup>
$s_{\text{cycle}}$	covered distance of a test cycle in km
$r_{\text{egr}}$	egr-rate
$s_{\text{egr}}, s_{\text{lp-egr}}$	position of the (low-pressure) egr valve
$s_{\text{sf}}$	position of the swirl flaps
$s_{\text{t}}$	position of the guide vanes of the turbocharger
$s_{\text{tv}}$	position of the throttle valve
$\text{soot}_{\text{limit}}$	emission standard limit for soot in mg/km
$\sigma^2, \hat{\sigma}^2$	(estimated) noise variance / variance of Gaussian weighting function
$\sigma_{\text{sensor}}$	series variations for sensors and construction tolerances
$T$	time constant in s / temperature in °C
$T_0$	sampling time in s
$T_{2i}$	intake temperature in °C
$t$	continuous time in s
$t_{\text{cycle}}$	required time for a test cycle in s
$\theta, \boldsymbol{\theta}$	parameter (vector)
$u, \mathbf{u}$	input (vector)
$u_{\text{inj}}$	fuel injection quantity in mm <sup>3</sup> /cyc
$\dot{v}_{\text{exh}}$	volume flow rate of the exhaust in m <sup>3</sup> /s

$w, \mathbf{w}$	parameter (vector)
$\mathbf{X}$	input matrix
$x, \mathbf{x}$	regressor (vector) of a polynomial
$x_{2i}$	gas composition at intake
$x_{\text{eng,out}}$	gas composition after combustion
$x_{\text{egr}}, x_{\text{lp-egr}}$	gas composition of the (low-pressure) egr
$y, \mathbf{y}$	output (vector)
$\hat{y}, \hat{\mathbf{y}}$	modelled output (vector)
$\mathbf{z}$	operation point / z-regressor of a multi-model structure

### Abbreviations

BSFC	Brake Specific Fuel Consumption
CARB	California Air Resource Board
CO	Carbon Monoxide
DoE	Design of Experiments
CFD	Computational Fluid Dynamics
ECU	Engine Control Unit
EGR	Exhaust Gas Recirculation
EPA	Environmental Protection Agency
FTP	Federal Test Procedure
HC	Hydrocarbons
i.i.d.	identical independent distributed
LEV	Low Emission Vehicle
LLM	Local Linear Model
LM	Local Model
LOLIMOT	Local Linear Model Tree
LOPOMOT	Local Polynomial Model Tree
LPM	Local Polynomial Model
LS	Least Squares
MISO	Multi Input Single Output
NEDC	New European Driving Cycle
$\text{NO}_x$	Nitrogen Oxides (NO and $\text{NO}_2$ )
PM	Particulate Matter
PRBS	Pseudo Random Binary Signal
RCP	Rapid Control Prototyping
RMSE	Root Mean Square Error
SULEV	Super Ultra Low Emission Vehicle
ULEV	Ultra Low Emission Vehicle

---

# Abstract

Modern Diesel engines require a model based optimisation of the engine control to fully exploit the additional degrees of freedom of modern engines. For identification of combustion engines, different experimental model structures are presented and compared to each other. The local adaptive model approach *LOPOMOT* is derived from the the local linear model approach *LOLIMOT* and an adaptive polynomial approach. Further regarded model structures are the in automotive industry well known look-up tables and the individual approximators kernel models. The model structures are generally presented and are rated with regard to applications in an electronic control unit.

For the identification of the combustion engine, the combustion outputs  $\text{NO}_x$ , soot and the engine torque are regarded. Experimental models are presented for measurements from the engine test bed. Stationary and dynamic effects are modelled separately, to avoid the influence of measurement dynamics. Thus, stationary measurements can be applied to identify the combustion models. The connection of these stationary combustion models to a dynamic air path model enables a dynamic overall simulation of the Diesel engine. The stationary and the dynamic model qualities are demonstrated using measurements from the engine test bed.

The models are then applied for a stationary and a dynamic optimisation of control functions for the engine control unit. At first a local optimisation is presented for the stationary optimisation, which shows the Pareto front of the emissions  $\text{NO}_x$  and soot. The subsequent global optimisation minimises the fuel consumption over a test cycle and formulates the emission limits as constraints. Initial values for the global optimisation are taken from the results of the local optimisation. Finally, a robust global optimisation is presented, which regards model uncertainties and variations due to series tolerances.

For the dynamic optimisation, the trajectories of the air path actuators are optimised for a typical acceleration event. Because of the high computationally effort, such an optimisation can not be performed during engine operation, but it enables conclusions about suitable control structures. Thereafter, a smoke limitation based on the soot model is presented. This model based smoke limitation requires no additional calibration effort, but the model parameters are difficult to interpret. Therefore, a simplification to an open loop control structure with look-up tables is shown, which enables a manual fine tuning of the maps.

This dissertation contributes to the model based optimisation of engine control functions and presents new modelling and optimisation approaches. Furthermore, new model structures are compared to the in automotive industry well known look-up tables and assets and drawbacks are discussed.

# Kurzfassung

Moderne Dieselmotoren erfordern eine modellbasierte Optimierung der Motorsteuerungen um die zusätzlichen Freiheitsgrade komplexer Motoren bestmöglich auszunutzen. Zur Modellierung eines Verbrennungsmotors werden verschiedene experimentelle Modellstrukturen vorgestellt und miteinander verglichen. Der lokal adaptive Modellierungsansatz *LOPOMOT* wird aus dem lokal linearen Modellansatz *LOLIMOT* und einem adaptiven Polynomansatz hergeleitet. Des Weiteren werden die in der Automobilindustrie weit verbreiteten Kennfelder und die für ihre universelle Approximation bekannten Kernschätzer betrachtet. Die Modellstrukturen werden allgemein vorgestellt und anhand von Beispielen bezüglich ihrer Einsatzfähigkeit im Motorsteuergerät bewertet.

Bei der Modellierung des Verbrennungsmotors werden die Verbrennungsausgänge  $\text{NO}_x$ , Ruß und das Motordrehmoment betrachtet. Es werden experimentelle Modelle anhand von Messungen vom Motorenprüfstand präsentiert. Um den Einfluss von Dynamiken bei der Emissionsmessung zu eliminieren, werden stationäre und dynamische Vorgänge separat modelliert. Dadurch können stationäre Messungen zur Identifikation der Verbrennungsmodelle verwendet werden. Die Verbindung dieser stationären Verbrennungsmodelle zu einem dynamischen Luftpfadmodell ermöglicht die dynamische Simulation des Dieselmotors. Die stationäre und dynamische Modellgüte wird anhand von Messungen vom Motorenprüfstand gezeigt.

Die Modelle werden dann für eine stationäre und eine dynamische Optimierung von Steuerungs- und Regelungsfunktionen für die Motorsteuerung verwendet. Bei der stationären Optimierung wird zunächst eine lokale Optimierung vorgestellt, welche die Paretofront der Emissionen  $\text{NO}_x$  und Ruß zeigt. Bei der anschließenden globalen Optimierung wird der Verbrauch über einen Testzyklus optimiert und die Emissionsgrenzen als Nebenbedingungen formuliert. Die Startwerte für die globale Optimierung stammen dabei von der lokalen Optimierung. Schließlich wird eine robuste globale Optimierung präsentiert, welche Modellunsicherheiten und Streuungen aufgrund von Serientoleranzen berücksichtigt.

Bei der dynamischen Optimierung werden die Trajektorien der Luftpfadaktoren für einen typischen Beschleunigungsvorgang optimiert. Eine solche Optimierung kann aufgrund des erheblichen Rechenaufwandes nicht während des dynamischen Motorbetriebs durchgeführt werden, ermöglicht jedoch Rückschlüsse über mögliche Steuerungs- und Regelungsstrukturen. Anschließend wird eine auf dem Rußmodell basierte Rauchbegrenzung vorgestellt. Die präsentierte modellbasierte Rauchbegrenzung erfordert keinen zusätzlichen Applikationsaufwand, die verwendeten Modellparameter sind jedoch schwierig zu interpretieren. Daher wird eine Vereinfachung zu einer Steuerungsstruktur mit Kennfelder präsentiert, welche für eine manuelle Feinjustierung verwendet werden kann.

Diese Dissertation leistet einen Beitrag zur modellbasierten Optimierung von Steuerungs- und Regelungsfunktionen und stellt neue Modellierungs- und Optimierungsansätze vor. Darüber hinaus werden neue Modellstrukturen mit der in der Automobilindustrie verbreiteten Kennfeldstrukturen verglichen und Vor- und Nachteile diskutiert.

---

# 1 Introduction

---

This thesis is about the model-based optimisation of the engine control. In this chapter, a motivation to utilise model-based approaches is given. Emissions standards are presented and challenges for an engine development are discussed. An overview of previous research on the subject is presented and a survey of this thesis is given.

---

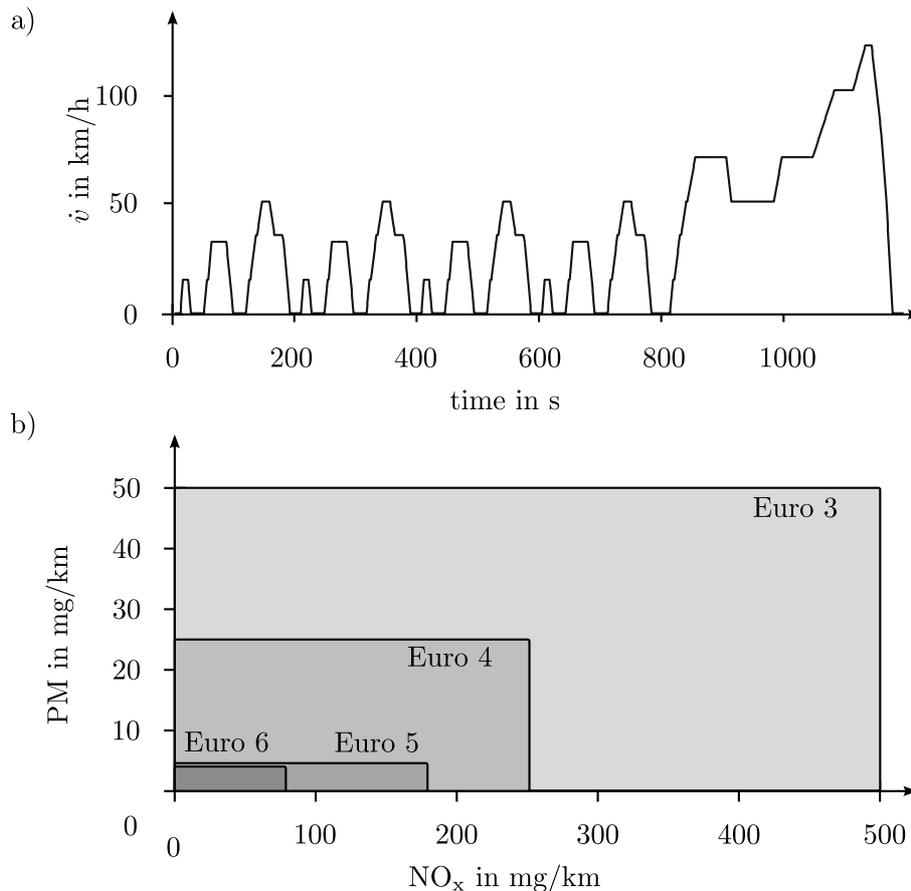
This thesis is a contribution to the model-based optimisation of combustion engines. Concrete applications are given with regard to an 1.9 l Common-Rail Diesel engine, but the presented approaches are transferable to other engines also with different fuels. In the following, a motivation for the model-based optimisation of the engine control is given and the objectives are discussed.

## 1.1 Model-based optimisation of the engine control

Fuel prices have increased significantly in recent years, which is why customers request more efficient cars. To comply this request, the combustion engine increases permanently in its complexity. Besides this, combustion engine emissions are ecologically damaging and have adverse effects on the human health. They are therefore limited by law and these limits are getting more and more stringent. The need to fulfil these limits further drives the complexity of combustion engines.

There are various emission standards for Europe, North America, South America, Asia and Australia with multiple differences in the countries. Limits are given for carbon monoxide CO, hydrocarbons HC, nitrogen oxide NO<sub>x</sub> and particulate matter PM [33]. The emission standards in Europe are defined by the *European Commission*. Emission limits are formulated over the *New European Driving Cycle NEDC*, depicted in Fig. 1.1 a). Classifications are given from *Euro 1* (became law in 1992) to *Euro 6* (becoming law in 2014) given in mg/km with regard to the NEDC. Further classifications, like *Euro 7*, are expected to be launched.

Emission standards in other countries differ in the limit values and also in the test cycles over which these limits are defined. The most stringent limits are given by the *California Air Resources Board CARB* [26], why these are a benchmark for other emission standards. The limits are defined over the *Federal Test Procedure FTP-75*. The emission standards are *Low Emission Vehicle LEV I* (became law in 1992) to *LEV III* (becoming law in 2015). Cars are further classified to *Low Emission Vehicle LEV*, *Ultra Low Emission Vehicle ULEV*, *Super Ultra Low Emission Vehicles SULEV* and others. Additionally, limits are defined over the lifetime of an engine. States in North America can either apply the emission standards of the CARB or the less stringent emission standards of the *Environmental Protection Agency EPA*. Emissions standards of the EPA are given as *Tier I* and



**Figure 1.1:** a) Velocity diagram of the *New European Driving cycle NEDC* over which the European emission standards are defined. b) Limit values for NO<sub>x</sub> and PM emissions for various European emission standards for Diesel engines.

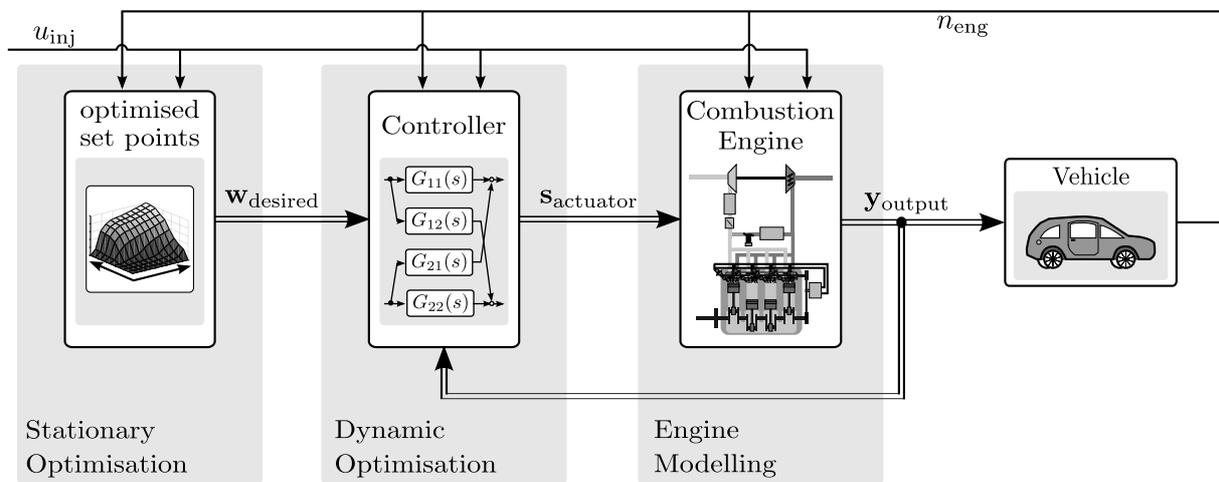
*Tier II*, which are again classified in several *bin*'s [32]. There are additional limits defined with regard to a combination of the test cycles FTP-75, US06 (highway cycle) and SC03 (cycle with air conditioning switched on).

The optimisation approaches presented in this work are with regard to the European standards, but they can directly be applied to alternative emission standards. The regarded emissions are NO<sub>x</sub> and PM. The emissions HC and CO are converted by an oxidation catalyst after light off temperature by a conversion rate above 90% [101, 105]. Hence, the optimisation of the HC and CO emissions mainly concerns the optimisation of the engine cold start. Since the engine cold start is not regarded in this work, it is focussed on the emissions NO<sub>x</sub> and PM. Particulate Matter are measured as soot, which is used as synonym in the following.

There is a trade-off between NO<sub>x</sub> and soot emissions, meaning that one emission increases if the other is lowered. In recent years, both emission limits have become more and more tightened, see Fig. 1.1 b), why the fulfilment of these limits is one of the main challenges in the calibration process of modern Diesel engine controls. For a complex engine it may not be possible to find the optimum manually by test bed experiments, why model-based approaches are required to fully exploit the

additional degrees of freedom of modern combustion engines. A manual optimisation may further require a long time for an engine calibration, which is, because of the shorter development cycles and the various calibrations for the different emission standards, not given. Furthermore, test bed time is expensive why long test runs must be avoided.

For a model-based optimisation, a model of the combustion engine is necessary. Depending on the engine complexity and the applied optimisation, various model structures are suited and their parameters are either identified with a theoretical or an experimental modelling approach. The determination of a combustion engine model is referred to as *engine modelling*. Given an engine model, the optimisation itself is divided into a *stationary optimisation* and a *dynamic optimisation*, see Fig. 1.2. Task of the stationary optimisation is the determination of stationary set points. These are then the desired values for the engine control, which is optimised by the dynamic optimisation.



**Figure 1.2:** Simplified control structure of a combustion engine. The set points are derived from a *stationary optimisation* and are the desired values for the engine controllers. The design and the parametrisation of these controllers are the objectives of the *dynamic optimisation*. To apply model-based approaches for these optimisations, a model of the combustion engine is necessary, which is obtained by *engine modelling*. All these blocks are influenced by the engine operation point, defined by the engine speed  $n_{\text{eng}}$  and the injection quantity  $u_{\text{inj}}$ .

Besides the optimisation of the engine control, the implementation in the *Engine Control Unit ECU* is a challenging task. ECUs are limited in computing and memory capacity and are further limited to real time capability. The increased engine complexity requires however more advanced controls for which several models need to be simulated on the ECU. The efficient implementation of these is therefore also a concern of the engine development process.

## 1.2 Previous work

### Stationary optimisation

An early model-based stationary optimisation approach, presented by [74], optimises the set points directly on the test bed, which is also referred to as *online optimisation*. Instead of applying a model for the input-output behaviour, the target function, which has to be minimised, is modelled by a second order polynomial model and minimised online on the test bed. This supports the test bed engineer during the manual optimisation on the test bed, but gives no further information about the process. The test bed engineer can further be supported by combining such an optimisation with an automation of the test bed [1]. Another approach that works online but applies models for the fuel consumption, the HC emissions and the knock limit is presented by [104]. It has the advantage that obtained process informations can also be utilised *offline*. Approaches that solely optimise the engine offline after data collection and modelling are presented by [48, 88]. Therefore, models for emissions and fuel consumption are identified at first on the engine test bed [150] and are then applied for a model-based optimisation on a desktop PC at the office.

The offline optimisation of the stationary engine control with experimental models is the focus of several works in recent years [87, 138, 184]. These optimisations regard either each engine operation point separately or several operation points simultaneously. The former is referred to as local optimisation and the latter as global optimisation. Overviews of local and global optimisation approaches are given in [27, 115, 149]. Applied loss functions are either multi-objective loss functions, regarding the emissions and the fuel consumption in one equation related by weights, or constrained loss functions, optimising the fuel consumption with regard to several constraints [149]. A combination of a local and a global optimisation is presented in [47]. There a local optimisation is performed in an inner loop and the fulfilment of emission standards is checked by a dynamic simulation of the test cycle in an outer loop. If the emission standards of the outer loop are violated, the local optimisation criterion is adjusted by the weights. The outer loop can therefore be regarded as global optimisation.

Besides the optimisation with experimental models, set points can also be derived from the optimisation of a complex fluid dynamical model [57, 77, 155]. Such models are also frequently applied to optimise the combustion process including the design of the combustion chamber [77, 101]. A commercial tool for a detailed engine simulation is GT-SUITE with its well-known software package GT-POWER [37]. An experimental possibility to optimise the combustion process is by visualisation of the flame front with optical sensors [58, 183].

### Dynamic optimisation

Dynamic optimisations mainly concern the design of closed loop controls. Common designs control the air path states, see e. g. [181], but in recent publications the direct control of emissions is suggested, see e. g. [169]. A detailed overview of closed loop controls is given in Sect. 5.3. Besides

closed loop controls, the design of feedforward controls is part of a dynamic optimisation. A model-based optimisation of the engine actuator trajectories for an acceleration event is presented in [16]. A similar optimisation on the test bed applying no model is shown by [6]. These optimisations of actuator trajectories require however the knowledge of the engine operation point trajectory and a significant computing time, why they are not feasible during engine operation.

Alternatively, the actuator control can be extended by a differential element, calibrated over the engine operation point, to account for the dynamics [46]. An even more flexible structure is suggested in [115], where an extra dimension is introduced for the degree of dynamic to determine dynamic set points. A similar approach suggested in [134] applies two types of set points, one for stationary operation and one for dynamic operation (smoke operation). A linear interpolation is used to switch between those set points. The design of a smoke limitation is a further part of the dynamic optimisation. A smoke limitation limits the injection quantity, e. g. by a limit value for the air to fuel ratio  $\lambda$  [124], to avoid visible soot emissions.

Modern ECUs can also shift a part of the dynamic optimisation into the ECU. The optimisation of the injection angle with regard to a loss function is proposed in [48]. Other modern controls apply model-based controls [55, 146, 181]. These are relatively easy to calibrate and can well be transferred from one engine to another, which is advantageous to reduce the calibration time. They require however a higher computing capacity.

### Engine modelling

To apply model-based approaches, models for the combustion engine are required. The methods to determine these models depend on the applied model structures. It can in general be differentiated between *crank angle synchronous models* and *mean value models*. The crank angle synchronous models can be categorised by their level of detail [101]. The mean value models are categorised into stationary or dynamic and into global or local. An overview of model structures is given in Chap. 3. For the identification of these models, either stationary or dynamic measurements are required. Various identification techniques for the identification of combustion engines are presented in [66, 189]. A stationary model structure requires stationary measurements, distributed over the drivability space. A common approach to distribute the measurements is by a D-optimal design [22, 137]. The generation of the D-optimal design is in [41] combined with the determination of the drivability space. The determination of the drivability space is also addressed in [133]. The chronological order of measurement points for a D-optimal design is discussed in [148].

A drawback of the stationary model identification is that for each measurement it has to be wait that the process is in steady-state. Thus, a major part of the measurements is unused. In contrast, a dynamic model structure is identified with dynamic measurements applying the entire measurement data. A dynamic excitation of the process can be given by a PRBS signal [67, 186]. An application of such a dynamic excitation for the identification of combustion engines is presented in [151]. From the identified dynamic models it is also possible to extract the stationary behaviour [85, 139, 149]. If the dynamics of the process are relatively fast, the stationary behaviour can also

be identified by quasistationary measurements [89, 139, 152]. The stationary measurements have however the advantage to average the noise effects, why these are more accurate than dynamic or quasistationary measurements. Hence, the type of measurement depends on the applied model structure and the required model accuracy.

Besides these general identification strategies, it can be advantageous to adjust the identification signal to the identified model structure. This is for example the case for a semi-physical model of the air path, where the various components are identified separately [109]. Another example is the identification of a torque model with injection angle sweeps, as described in [20, 24, 173]. A similar approach for the identification of a torque model with a non-linear optimisation algorithm is presented in [144].

Various of these modelling and optimisation approaches are integrated in several commercial tools. These tools support the engineers at the engine development process either online at the test bed or offline in the office. Common tools are among others *ASCMO* by ETAS [86], *CAMEO* by AVL [19], *TOPexpert Suite* by FEV [144] and *Easy DoE ToolSuite* by IAV [14].

### 1.3 Thesis outline and new contributions

The objective of this thesis is the integrative presentation and the further development of model-based optimisation techniques for combustion engines. It focuses on the experimental modelling of the combustion outputs  $\text{NO}_x$ , soot and the engine torque, and the optimisation of these with regard to a given emission standard and the fuel consumption. To identify combustion models, the modelling toolbox LOLIMOT is further developed and dynamic and stationary effects of the engine are separated. Local and global stationary optimisations of the engine control are presented and influences of model uncertainties and series variations are regarded. A dynamic optimisation is shown for a typical acceleration event and a model-based smoke limitation is introduced. The implementation of the results in the ECU is regarded, for what different model structures are compared to the common used look-up tables. The main contributions can be summarised to:

- Further development of the modelling toolbox LOLIMOT to local adaptive polynomial models.
- Experimental mean value models for the combustion outputs  $\text{NO}_x$ , soot and the engine torque with a separation of stationary and dynamic effects.
- Local and global stationary optimisations of the engine control with taking account of model uncertainties and series variations.
- Design of a model-based smoke limitation based on the presented soot model.
- Comparison of alternative model structures to look-up tables with regard to an implementation in the ECU.

In *chapter 2* five experimental model structures are presented: look-up tables, polynomial models, LOLIMOT, LOPOMOT and a Kernel model. For look-up tables a regularisation is introduced and possible combinations of several two-dimensional look-up tables are presented to map higher dimensional relations. Polynomial models are derived from the Taylor series of a function and a selection algorithm for significant regressors is introduced. The multi-model structure LOLIMOT is described and it is further developed to the new toolbox LOPOMOT, which combines LOLIMOT with the previously presented adaptive polynomial models. A comparison using two test functions shows the advantages of these further developments. Finally, the Kernel model structure is presented.

In *chapter 3* mean value combustion models for the outputs  $\text{NO}_x$ , soot and the engine torque are introduced. Several output transformations and various model inputs are discussed. The applied stationary global-local model structure and the measurement design is shown. Training and validation results are presented for stationary and dynamic measurements. An intersection plot tool to analyse the stationary model behaviour is presented and the  $\text{NO}_x$ -soot trade-off is discussed. The dynamic formation of emissions is analysed for several engine transients and the influences of measurement dynamics are shown.

In *chapter 4* the stationary optimisation of the engine control is presented. Objective functions for a local and a global optimisation are discussed and derived from a given emission standard with its defined test cycle. The influences of the applied optimisation algorithm, the intake temperature, the model uncertainties and the series variations are shown for the local optimisation. The model uncertainties and series variations are integrated in a robust global optimisation such that obtained optimisation results are non-sensitive to these variations and uncertainties. The optimisation results are discussed with regard to a stationary, a quasistationary and a dynamic simulation of the NEDC.

The dynamic optimisation of the engine control is addressed in *chapter 5*. An objective function is derived from the stationary optimised values and an optimisation of the air path actuator trajectories for a typical acceleration event is presented. Various closed loop controls are discussed and possible control variables and control actuators are listed. The advantage of a coupled control actuator is presented by an example. A model-based smoke limitation is introduced to avoid visible smoke during acceleration events. The smoke-limitation is based on the soot model and requires no additional calibration effort. The performance is shown for measurements on the test bed.

The various experimental model structures presented in Chap. 2 are compared to each other with regard to an implementation in the ECU in *chapter 6*. Evaluation criteria are the model accuracy, the computing time and the required memory. Conclusions about the suitability of alternative model structures to the common used look-up tables are given. Thereafter, a torque model consisting of several two-dimensional look-up tables is presented. The inversion of this model for the control of the injection quantity is discussed. Finally, an open loop control of the smoke limitation based on look-up tables is derived from the model-based smoke limitation presented in Chap. 5.

The presented methods and results are summarised in *chapter 7*. An outlook for further research is given. Several additional information can be found in the appendix.

---

## 2 Model Structures

---

Five experimental model structures are presented in this chapter: look-up tables, polynomial models, LOLIMOT, LOPOMOT and a Kernel model. A regularisation is presented for the look-up tables to identify weak excited parameters and to smooth a look-up table. Combinations of two-dimensional look-up tables to higher dimensional structures are shown and the inversion of a look-up table is discussed. For the polynomial models, a selection algorithm is presented to select the significant regressors of a Taylor series. Then, the multi-model structure LOLIMOT is described and combined with the local adaptive polynomial models to the new modelling toolbox LOPOMOT. Comparisons to LOLIMOT are given using two test functions. Finally, the data driven Kernel model is presented.

---

The calibration of a combustion engine with regard to the many different laws requires an engine modelling. It is generally advisable to apply physical models whenever possible. However, some processes, such as the emission formation, are only partially known or require an unreasonable effort to be modelled physically. Furthermore, the derived physical models might be computationally too intensive for a considered application. In these cases experimental models, which are often referred to as *black-box* models, can be applied. For black-box models no a-priori knowledge is necessary, in contrast to the physically based *white-box* [92]. Since the transfer between white and black box models is fuzzy, there are different shades of grey possible depending on how much a-priori knowledge can be included into the model structure [64, 67, 93, 94]. The here regarded models are experimental models with fixed model structures and are therefore regarded as *dark-grey / steel-grey-box* models if a-priori knowledge is included or as black-box models if not. Overviews of experimental models for linear and non-linear static and dynamic systems are given in [67, 72, 91, 92, 118, 157].

The presented model structures are *look-up tables*, *polynomial models*, *LOLIMOT*, *LOPOMOT* and a *Kernel model* structure. Look-up tables are introduced in Sect 2.1. Because of their universal approximation quality, their fast computation and their good interpretability, look-up tables are commonly used model structures for automotive applications. At first, the identification of look-up tables is presented applying a least-squares approach, a recursive least squares approach and an approach with intermediate models. Then, a regularisation is presented that allows to identify not excited parameters and enables to smooth a look-up table in dependence on a regularisation parameter. Since look-up tables are usually limited to two input variables, combinations of two-dimensional look-up tables are introduced to map higher dimensional relations. Finally the inversion of a look-up table is regarded.

In Sect. 2.2 polynomial models are motivated by the Taylor series of a continuous function. Polynomials can approximate any continuous function locally arbitrary well, but the number of regressors

increases exponentially with an increase of the input dimension and the polynomial order. Therefore, a selection of significant regressors is presented, consisting of a criterion of fit and a heuristic selection algorithm. This selection enables an adaption of the polynomial model to the process. Based on this, the so called *Design of Experiments* is motivated.

The LOLIMOT structure is shortly introduced in Sect. 2.3. It is a multi-model approach, which composes local affine models by means of Gaussian weighting functions to a global model output. The partitions on which the local models are valid are derived from a tree construction algorithm applying axis orthogonal splits.

Based on LOLIMOT and the adaptive polynomial models presented before, the LOPOMOT structure is derived in Sect. 2.4. In contrast to LOLIMOT, local polynomial models are utilised. Thus, the good local approximation quality of polynomials is combined with the flexibility of a multi-model approach. The selection of regressors adapts the local polynomial models to the local process behaviour and improves the overall model quality. Based on the applied local selection criterion, a global criterion of fit is derived to select the best model partition. The number of applied partitions is in general less compared to LOLIMOT, since the local models are nonlinear. A comparison between LOLIMOT and LOPOMOT is given using a static non-linear and a Hammerstein test model simulation.

Finally, the data driven Kernel model is introduced in Sect. 2.5. The Kernel model mainly consists of the measured training data. For an evaluation of the Kernel model, all measured output values from training data are weighted and summed up to a global model output. The weights are determined by the kernel function, which depends on the measured inputs from training data and a bandwidth parameter.

The presented model structures have different properties, why the suitability of a model structure depends on the field of application. Models can either be used to map functional relations offline on a desktop PC or online on an ECU during engine operation. The demands for models evaluated online on an ECU are in general much more stringent, since these models need to be calculated in real time, while having a limited amount of computing and memory resources. A comparison of the presented model structures with regard to computing time, model quality, required memory and interpretability is given in Chap. 6.

## 2.1 Look-up tables

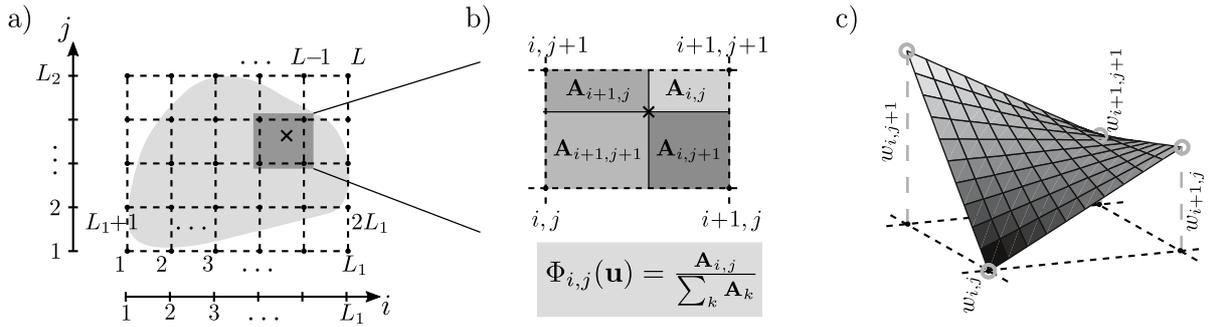
Look-up tables, also called characteristic maps, are commonly used model structures for automotive applications [105, 136]. They are defined by a grid over the input space, at which each grid point is assigned a grid point height. Thus, any functional relation can be mapped, if the grid points are distributed dense enough. A general parametric formulation is given by

$$f_{\text{look-up}}(\mathbf{u}) = \sum_{i=1}^L \Phi_i(\mathbf{u}) w_i, \quad (2.1)$$

where  $\Phi_i(\mathbf{u})$  is the weighting function for the  $i$ -th grid point and  $w_i$  is the corresponding grid point height.  $\mathbf{u}$  is the input vector and its dimension depends on the number of influencing variables. For a one dimensional dependency,  $\mathbf{u}$  is a scalar. Then the look-up table is a one-dimensional characteristic curve and the weighting function is given as a linear interpolation. The most common look-up tables for automotive applications are two-dimensional,  $\mathbf{u}(i) = (u_1(i), u_2(i))$ , with grid points distributed over a lattice with  $L_1$  partitions in  $u_1$  and  $L_2$  partitions in  $u_2$ . The distribution can either be equidistant or non-equidistant. In both cases the model can be written intuitively as

$$f_{\text{look-up}}(u_1, u_2) = \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} \Phi_{i,j}(u_1, u_2) w_{i,j}, \quad (2.2)$$

where  $\Phi_{i,j}(u_1, u_2)$  is the weighting function for the  $i$ -th grid point in the dimension of  $u_1$  and the  $j$ -th grid point in the dimension of  $u_2$ . The corresponding grid point height is denoted by  $w_{i,j}$ . Eq. (2.2) can also be written in the general form of eq. (2.1), by numbering the grid points from  $i = 1$  to  $L$  with  $L = L_1 L_2$ . An equidistant lattice distribution of the grid points and a possible numeration, starting with the grid point in the lower left corner of the lattice, is depicted in Fig. 2.1 a).



**Figure 2.1:** a) Lattice structure of a look-up table with equidistant distribution of the grid points. The rectangular input space covers the model input space (light grey area). b) Calculation of the grid point weights for one grid element by the bilinear interpolation, shown as area interpolation. c) Output of the bilinear interpolation for one grid element. The bilinear interpolation is a curved plane, if the four grid point heights  $\mathbf{w}$  are not located on a plane.

Instead of calculating all  $L$  weights in eq. (2.2), it is sufficient to regard the four grid points surrounding the model input  $\mathbf{u} = (u_1, u_2)$ . The model input is indicated by the cross in Fig. 2.1 a) and the surrounding grid points are covered by the dark grey area. The surrounding grid points can be determined by several comparison operations. Efficient algorithms for this can be found in [145], why it is no further discussed here. Given the surrounding grid points, the weights are computed by a weighting function  $\Phi$ , which is a bilinear interpolation. This bilinear interpolation is equivalent to an area interpolation as sketched in Fig. 2.1 b), see [62]. The four weights are then determined

by

$$\Phi_{i,j}(u_1, u_2) = \frac{1}{A} \begin{cases} A_{i+1,j+1} = |u_{1,i+1} - u_1| |u_{2,j+1} - u_2| & : \text{lower left corner} \\ A_{i,j+1} = |u_{1,i} - u_1| |u_{2,j+1} - u_2| & : \text{lower right corner} \\ A_{i+1,j} = |u_{1,i+1} - u_1| |u_{2,j} - u_2| & : \text{upper left corner} \\ A_{i,j} = |u_{1,i} - u_1| |u_{2,j} - u_2| & : \text{upper right corner} \end{cases} \quad (2.3a)$$

$$\text{with} \quad A = |u_{1,i+1} - u_{1,i}| |u_{2,j+1} - u_{2,j}|. \quad (2.3b)$$

The coordinates  $u_{1,i}$  and  $u_{1,i+1}$  are the lower respectively upper coordinate of the surrounding grid points of  $u_1$  and the coordinates  $u_{2,j}$  and  $u_{2,j+1}$  are the analogue coordinates in the second dimension  $u_2$ .

Given the four weights, they are multiplied by their corresponding grid point heights and summed up to a look-up table output, see eq. (2.2). A continuous evaluation of the bilinear interpolation over one grid element is depicted in Fig. 2.1 c). The grid point heights are indicated as grey circles. It can be seen that the continuous evaluation in between the four surrounding grid points is a curved plane, if the four surrounding grid point heights are not located on a plane.

### 2.1.1 Identification

The identification of a look-up table mainly involves the determination of the grid point heights. According to [168] there are three possibilities to identify a look-up table. The *first* is to measure directly on the grid points and employ the measured values as grid point heights. The *second* possibility is more flexible as the measured values can be distributed arbitrarily in the input space. The heights are then calculated by an identification algorithm. For both, the first and the second possibility, the grid structure is fixed. For the *third* possibility the grid structure is optimised simultaneously to the grid point heights. Therefore, a linear inner loop optimisation of the grid point heights and a non-linear outer loop optimisation of the grid structure is introduced in [120].

In the following it is assumed that the grid structure is fixed and that the measurements are distributed arbitrarily in the input space, which relates to the second possibility mentioned. To identify the grid point heights, a least-squares approach, a recursive least-squares approach and an approach with an intermediate model are presented. The recursive least-squares algorithm enables an adaptation of the look-up table during an application and the identification with an intermediate model enables the reduction of a complex model structure to a look-up table. An intermediate model can also be applied to introduce a regularisation.

### Least squares estimation

It is assumed that  $N$  measurements  $(\mathbf{u}(i), y(i))$  are distributed over the input space. Applying eq. (2.1) for these  $N$  measurements a linear set of equations can be formulated as

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{e}, \quad (2.4)$$

with the input matrix and the output vector

$$\mathbf{X} = \begin{pmatrix} \Phi_1(\mathbf{u}(1)) & \Phi_2(\mathbf{u}(1)) & \dots & \Phi_L(\mathbf{u}(1)) \\ \Phi_1(\mathbf{u}(2)) & \Phi_2(\mathbf{u}(2)) & \dots & \Phi_L(\mathbf{u}(2)) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_1(\mathbf{u}(N)) & \Phi_2(\mathbf{u}(N)) & \dots & \Phi_L(\mathbf{u}(N)) \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{pmatrix}, \quad (2.5)$$

the unknown grid point heights

$$\mathbf{w}^T = (w_1, w_2, \dots, w_L) \quad (2.6)$$

and the measurement error  $\mathbf{e}$ . The weighting function  $\Phi_j(\mathbf{u}(i))$  determines the influence of the measurement point  $\mathbf{u}(i)$  on the  $j$ -th grid point and is known from eq. (2.3) for the given inputs. For the regarded two dimensional case, each measurement influences a maximum of four grid points. Hence, for each row in  $\mathbf{X}$ , a maximum of four weights are unequal to zero. To determine all grid point heights, the inputs need to be well distributed over the lattice structure and more measurements than unknowns are required,  $N \geq L$ . Minimising the sum of squared errors

$$J = \mathbf{e}^T \mathbf{e}, \quad \text{with} \quad \mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{w}, \quad (2.7)$$

leads without any assumption about the measurement error distribution to the least squares solution, see e. g. [34, 35, 50, 67, 92],

$$\hat{\mathbf{w}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (2.8)$$

An often applied assumption is that the measurement error  $\mathbf{e}$  is identical independent distributed (i.i.d.). With  $\mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma^2)$ , where  $\mathcal{N}$  is the normal distribution, eq. (2.8) is the best linear unbiased estimator (*BLUE*), which means that the estimate has the minimal variance among all linear unbiased estimators, see e. g. [34, 92]. Such an estimate is also denoted as Markov estimate, see e. g. [50, 35, 67]. If the inputs are not sufficiently distributed in the input space, several grid points might not be excited and need therefore be excluded for identification. Alternatively, a regularisation can be introduced as is shown in the following section. Such a regularisation reduces the number of effective parameters and smooths the look-up table.

### Recursive least squares estimation

For the recursive least squares approach it is assumed that measurements are added to the model when the model is already partially identified. Thus, the model can adapt to the process, which

might e. g. be necessary to compensate drift or ageing effects. For the recursive least squares solution, as presented in [67], the parameters are updated with each new measurement by

$$\hat{\mathbf{w}}(k + 1) = \hat{\mathbf{w}}(k) + \gamma(k + 1) (y(k + 1) - \hat{y}(k + 1)). \quad (2.9)$$

Hence, the parameter vector at discrete time  $k + 1$ ,  $\hat{\mathbf{w}}(k + 1)$ , is given by the parameter vector at time  $t$ ,  $\hat{\mathbf{w}}(k)$ , being adjusted by the weighted difference between the new measurement  $y(k + 1)$  and the predicted output  $\hat{y}(k + 1)$ . Details about the calculation of the adaption weight  $\gamma(k + 1)$  and about the selection of initial values,  $\mathbf{w}(0)$  and  $\gamma(0)$ , are given in App. D.1. Another detailed discussion about recursive least squares algorithms for look-up tables with simplifications for an accelerated computation is given in [180].

### Estimation with intermediate models

Besides these direct linear identification techniques, look-up tables can also be identified indirectly using intermediate models [163, 167]. As intermediate model, neuronal networks  $\hat{f}_{\text{nn}}$ , such as multilayer perceptrons or radial basis function networks, can be applied. These are known as universal approximators [61], but they are not suited for an implementation on an ECU, as they are computationally intensive and difficult to interpret. However, if such a complex model is given, it can be evaluated on the grid points and stored as look-up table,

$$\hat{w}_i = \hat{f}_{\text{nn}}(\mathbf{u}_i). \quad (2.10)$$

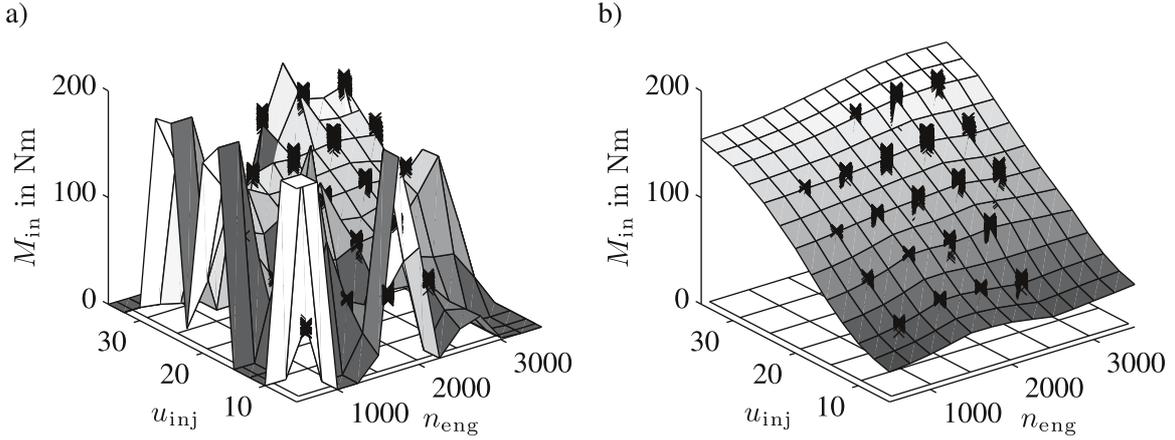
Advantageous of the identification with an intermediate model is its implicit regularisation, provided that the intermediate model possesses less parameters than the look-up table. Such an identification of look-up tables is also suited to simplify a complex model, e. g. a detailed physical model, to a look-up table.

#### 2.1.2 Regularisation

Depending on the process and the measurement design, there might be regions in the look-up table where no or only few measurements exist. Then the corresponding grid points are not or only weakly excited. This lack of excitation is either due to not measurable grid points or due to a poor measurement distribution. Not reachable grid points are often in the boundary region of a look-up table, since the model input space is in general not equal to the rectangular input space of a look-up table, see the light grey area in Fig. 2.1 a). If grid points are not excited, the corresponding heights can not be estimated. If grid points are weakly excited, which means that only few measurements exist in their neighbourhood, the input matrix  $\mathbf{X}$  is ill-conditioned and sharp differences in the grid point heights can occur.

The negative effects of a poor measurement distribution are shown in Fig. 2.2 a). The plot shows the identified two-dimensional base look-up table for the torque model discussed in more detail in

Sect. 6.2. Measurements are indicated as black crosses and the lattice structure of the grid is depicted on the ground. If measurements are not directly located on the grid points, sharp differences occur as can especially be seen for low engine speeds  $n_{\text{eng}} < 1000$  rpm. Furthermore, there are grid points in the boundary region that are not excited and set to zero, see e. g. high engine speeds with low injection quantities. To avoid these sharp differences and to enable the estimation of not excited grid points, a regularisation is introduced.



**Figure 2.2:** a) Identified base look-up table for the inner torque of a 1.9 l Common-Rail Diesel engine without regularisation and b) with regularisation presented in more detail in Sect. 6.2. The measurement points applied for model training are plotted as crosses. Sharp differences occur due to not excited grid points or due to a poor measurement distribution. The grid structure is indicated by the lattice in the  $n_{\text{eng}}, u_{\text{inj}}$ -plane. The engine speed  $n_{\text{eng}}$  and the injection quantity  $u_{\text{inj}}$  are given in rpm and  $\text{mm}^3/\text{cyc}$  respectively.

To determine a regularisation, the differences of grid point heights to neighbouring grid point heights are penalised. Therefore, several rows are appended to the input matrix  $\mathbf{X}$  and the output vector  $\mathbf{y}$  of eq. (2.5), by

$$\mathbf{X}_{\text{reg}} = \begin{pmatrix} \mathbf{X} \\ \lambda \mathbf{\Gamma} \end{pmatrix} \quad \text{and} \quad \mathbf{y}_{\text{reg}} = \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix}, \quad (2.11)$$

with the additional rows

$$\lambda \mathbf{\Gamma} = \lambda \begin{pmatrix} -2 & 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots \\ 1 & -3 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots \\ 0 & 1 & -3 & 1 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \dots & \ddots & \ddots & \ddots & \dots & \ddots & \ddots & \ddots & \dots \\ 0 & 1 & 0 & 0 & \dots & 1 & -4 & 1 & \dots & 0 & 1 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \dots & \ddots & \ddots & \ddots & \dots & \ddots & \ddots & \ddots & \dots \end{pmatrix} \quad (2.12)$$

and the regularisation parameter

$$\lambda \in [0, \infty]. \quad (2.13)$$

For a two-dimensional look-up table, each grid point has two, three or four neighbours, depending on the location of the regarded grid point. A grid point in the corner has two neighbours, a grid point on the edge three and a grid point in the interior of the lattice four neighbours, see Fig. 2.1 a). Eq. (2.12) shows the regularisation for one grid point in the corner, then two grid points on the edge and further below one grid point in the interior of the lattice. Applying this regularisation, the identification of grid point heights remains a linear problem, for what the solution is given analogue to eq. (2.8) by

$$\hat{\mathbf{w}}_{\text{reg}} = (\mathbf{X}_{\text{reg}}^T \mathbf{X}_{\text{reg}})^{-1} \mathbf{X}_{\text{reg}}^T \mathbf{y}_{\text{reg}}, \quad (2.14)$$

or by substituting eq. (2.11) in eq. (2.14) this leads to the solution

$$\hat{\mathbf{w}}_{\text{reg}} = (\mathbf{X}^T \mathbf{X} + \lambda^2 \mathbf{\Gamma}^T \mathbf{\Gamma})^{-1} \mathbf{X}^T \mathbf{y}. \quad (2.15)$$

Hence, this can also be regarded as Tikhonov regularisation [164]. The degree of regularisation depends on the parameter  $\lambda$ . The larger  $\lambda$  is chosen the smoother the look-up table and  $\lambda = 0$  corresponds to the non-regularised solution. The characteristic of the regularisation is determined by the matrix  $\mathbf{\Gamma}$ . It is shown in App. D.2 that the presented intuitive penalisation of differences in grid point heights, see eq. (2.12), corresponds to a penalisation of the second derivatives for interior grid points and a penalisation of the first derivatives for grid points on the boundary. Furthermore, the penalisation for non-equidistant grid point distributions and an application of individual regularisation parameters for each dimension,  $\lambda_1$  and  $\lambda_2$ , are presented in App. D.2. Given the Tikhonov regularisation as in eq. (2.15), there are alternative penalisation possible, such as the ridge regression by choosing  $\mathbf{\Gamma} = \mathbf{I}$ , see [60, 99].

The performance of the introduced regularisation is shown in Fig. 2.2 b). The sharp differences of the not regularised look-up table are smoothed and regions where no measurements are available are well estimated. The suited regularisation parameter  $\lambda$  is determined by testing several regularisation parameters and selecting the best parameter value by a visual decision. The best regularisation parameter can also be determined analytically if an independent test set is available.

The size of a suited regularisation parameter  $\lambda$  depends on the process and the regarded application. Hence,  $\lambda$  needs to be determined individually for each process, for what several regularisation parameters need to be tested. This is uncomplicated from a computational point of view, since the regularisation of eq. (2.11) leads to a linear solution, eq. (2.15). Unfavourable is however that the suited regularisation parameter  $\lambda$  depends also on the number of measurements, since the influence of the additional rows in  $\mathbf{X}_{\text{reg}}$  decreases with an increase of the number of measurements. The regularisation parameter can therefore be determined by selecting the measurement independent regularisation parameter  $\tilde{\lambda} \in [0, \infty]$  and calculating  $\lambda$  by [129]

$$\lambda = \tilde{\lambda} \sqrt{\frac{\text{trace}(\mathbf{X}^T \mathbf{X})}{\text{trace}(\mathbf{\Gamma}^T \mathbf{\Gamma})}}. \quad (2.16)$$

This relation can also be utilised to derive a regularisation for the recursive least squares algorithm, see App. D.2. A regularisation for a recursive adapted look-up table is also necessary, since a look-up table would otherwise loose its smoothness with the time.

### 2.1.3 Combination of two-dimensional look-up tables

The presented identification and regularisation can easily be extended to look-up tables with more than the common two input variables. Problematic is however the memory demand for look-up tables with more than two inputs, due to the exponentially growing of number of grid points, also known as *curse of dimensionality* [15]. Approaches that avoid an exponential growing utilise hierarchical grid structures [145, 165, 166]. There, grid points are placed in the relevant regions rather than increasing the discretisation for the entire lattice structure. [166] concludes that thus look-up tables with up to five inputs are feasible. A hierarchical structure requires however a higher computing time, since the determination of grid point weights is more complex. Furthermore, the interpretability of look-up tables with more than two inputs is complicated.

Alternative to look-up tables with more than two inputs, several two-dimensional look-up tables can be combined to map higher dimensional relations. In [172] a torque model for gasoline engines is presented, consisting of a two-dimensional base look-up table being multiplied by the output of several correction curves. The base look-up table models the maximum torque and the several correction curves model efficiency factors. Alternative to this multiplicative combination, look-up tables can also be added, see [81]. A further possibility is to combine a complex model structure with a look-up table structure, e. g. [80]. The various combinations of look-up tables are presented in the following with regard to the general four dimensional function

$$y = f(u_1, u_2, u_3, u_4). \quad (2.17)$$

Another overview of various look-up table combinations is also given in [171].

#### Combination with a complex model structure

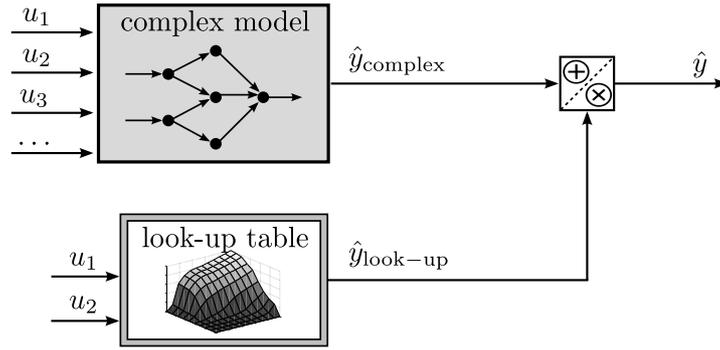
The combination of a look-up table with a complex model structure is shown in Fig. 2.3. Such a combination can for example be applied to model objective and subjective informations individually. Then the complex model is applied for the objective informations, while the look-up table is applied for the subjective informations [80]. Objective informations can e. g. be derived from physical laws or from the optimisation of an objective function. Subjective informations can be given by correction terms or by an individual tuning.

Another possibility is to utilise the look-up table as an adaptive model [82, 111]. The adaptive model is then used to model alterations of the process, while the complex model is applied for the basic behaviour. Thus, limitations for adjustments can easily be formulated for the look-up table such that malfunctions due to faulty corrections can be avoided.

The connection between the complex model and the look-up table is additive or multiplicative. Assuming an additive connection and that the look-up table depends on the inputs  $u_1$  and  $u_2$ , this structure can be written as

$$\hat{y} = \hat{y}_{\text{complex}} + \hat{y}_{\text{look-up}} \quad (2.18a)$$

$$= \hat{f}_{\text{complex}}(u_1, u_2, u_3, u_4) + \hat{f}_{\text{look-up}}(u_1, u_2). \quad (2.18b)$$

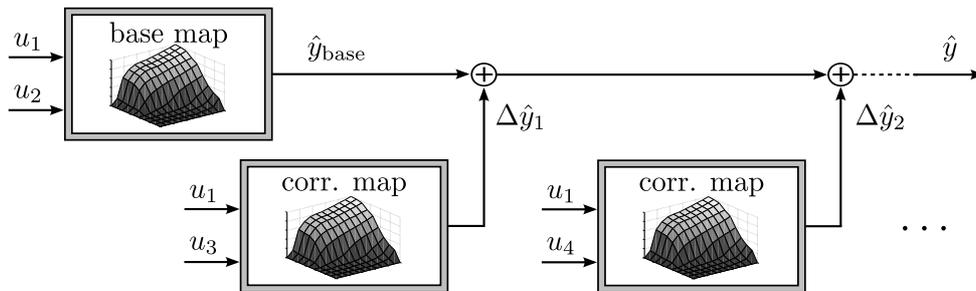


**Figure 2.3:** Combination of a complex model structure with a look-up table structure. The complex model is utilised to model the objective behaviour, while the look-up table is applied for subjective model tuning or as correction model. The combination of these models is additive or multiplicative.

The formulation with a multiplicative connection is analogue. Note that the model structure of the complex model is arbitrary for such a combination.

### Additive structure

For the additive structure, several two-dimensional look-up tables are added to map a higher dimensional relation, see Fig. 2.4. Each look-up table depends on a different input combination. It is however possible to apply an input variable for several look-up tables, here  $u_1$  for the base and the correction look-up tables. For  $p$  inputs and two-dimensional look-up tables there are  $\binom{p}{2}$  possible



**Figure 2.4:** Additive look-up table structure. Correction look-up tables are added to the base look-up table until a desired model quality is reached. The multiplicative structure is analogue with multiplications instead of additions.

input combinations. Hence, a maximum of  $\binom{p}{2}$  look-up tables can be applied for this additive structure. The number of required look-up tables depends on the process and the desired accuracy. The equation for the additive structure with three look-up tables, as shown in Fig. 2.4, is

$$\hat{y} = \hat{y}_{\text{base}} + \Delta \hat{y}_1 + \Delta \hat{y}_2 \quad (2.19a)$$

$$= \hat{f}_{\text{base}}(u_1, u_2) + \hat{f}_{\text{corr } 1}(u_1, u_3) + \hat{f}_{\text{corr } 2}(u_1, u_4). \quad (2.19b)$$

The parameters of the look-up tables can then either be determined by a) a *separate* identification of the look-up tables with several datasets or by b) a *simultaneous* identification of all look-up tables with one dataset.

- a) For the *separate* identification, the inputs of one look-up table are excited, while the other model inputs are kept constant. For the base look-up table the excited inputs are  $u_1$  and  $u_2$  and the constant inputs are  $u_3$  and  $u_4$ . Given measurements exciting  $u_1$  and  $u_2$ , the base look-up table can be identified with the approaches presented in see Sect. 2.1.1 and Sect. 2.1.2. Thereafter, the first correction look-up table is identified. Therefore, the inputs  $u_1$  and  $u_3$  are excited by another dataset, while  $u_2$  and  $u_4$  are kept constant. The correction look-up table is then identified with the residuals

$$\Delta y_1 = y - \hat{f}_{\text{base}}(u_1, u_2). \quad (2.20)$$

Further correction look-up tables, here  $\hat{f}_{\text{corr } 2}$ , are identified analogue.

- b) For the *simultaneous* identification, all inputs are excited within one dataset. With this dataset, at first a base look-up table  $f_{\text{base}}$  is identified. This base look-up table approximates  $f(u_1, u_2, u_3, u_4)$  in eq. (2.17) by a regression, here with regard to  $u_1$  and  $u_2$ . Given this base look-up table, the residual as in eq. (2.20) are calculated. These residuals are then applied for a regression with regard to the inputs of the first correction look-up table, here  $u_1$  and  $u_3$ . Again, further correction look-up tables are identified analogue.

The second approach enables the identification of all look-up tables with one dataset, whereas several datasets are necessary for the first approach. The first approach can therefore be utilised to calibrate an existing structure with systematic measurements, whereas the second approach is suited for a fast calibration of a higher dimensional model structure by a simultaneous excitation of all model inputs.

### Multiplicative structure

The multiplicative structure can be determined analogue to the additive structure. The approaches are identical, except for the determination of the residuals. The counterpart of eq.(2.20) is given by

$$\Delta y_1 = \frac{y}{f_{\text{base}}(u_1, u_2)}. \quad (2.21)$$

### Nested structures

Besides these parallel model structures, nested model structures can be applied. There are two possible nested model structures. For the first approach, the output of a look-up table is again introduced in another look-up table, whose output is again introduced in another look-up table and

so on, see Fig. 2.5. Thus, the model output is determined by the last correction look-up table, here given as

$$\hat{y} = \hat{f}_{\text{corr } 2}(\hat{y}_2, u_4). \quad (2.22)$$

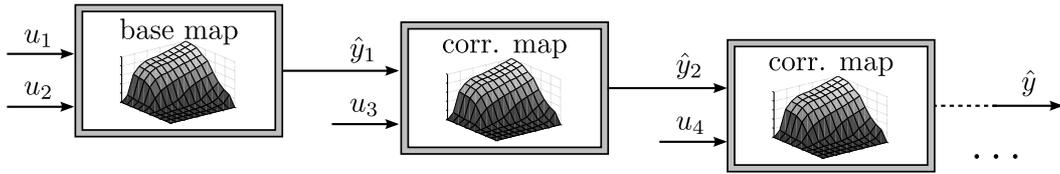
The output  $\hat{y}_2$  is again defined by a look-up table depending on an input, here  $u_3$ , and the output of the preceding look-up table, here  $\hat{y}_1$ . This is repeated until the base look-up table is reached such that

$$\hat{y}_2 = \hat{f}_{\text{corr } 1}(\hat{y}_1, u_3) \quad \text{and} \quad \hat{y}_1 = \hat{f}_{\text{base}}(u_1, u_2). \quad (2.23)$$

Substituting eq. (2.23) into eq. (2.22) gives the equation for the nested model structure

$$\hat{y} = \hat{f}_{\text{corr } 2}(\hat{f}_{\text{corr } 1}(\hat{f}_{\text{base}}(u_1, u_2), u_3), u_4). \quad (2.24)$$

Such a nested model structure can be advantageous if inner states exist for the intermediate values,  $\hat{y}_1$  and  $\hat{y}_2$ . If these states are measurable, each look-up table can be identified separately. However, if there is no physical interpretation for the intermediate values, this structure is hard to interpret and the identification is complicated.



**Figure 2.5:** Nested model structure. The output of a look-up table forms the input of the subsequent look-up table. The outputs  $\hat{y}_1, \hat{y}_2, \dots$  are therefore discretised. They should describe physical variables, such as inner process states, to be interpretable.

The second possible nested structure defines on each grid point of the base look-up table the parameters for another local look-up table, see Fig. 2.6. This can be interpreted as *global-local* model structure and is especially then suited, if the global inputs, here  $u_1$  and  $u_2$ , are selected as operation point (e. g. the engine speed  $n_{\text{eng}}$  and the injection quantity  $u_{\text{inj}}$  for a Diesel engine). The equation for the global look-up table is given analogue to eq. (2.1) by

$$\hat{y} = \sum_{i=1}^L \Phi_i(u_1, u_2) \hat{y}_i. \quad (2.25)$$

In contrast to eq. (2.1), the grid point heights  $\hat{y}_i$  are not constants but determined again by local look-up tables

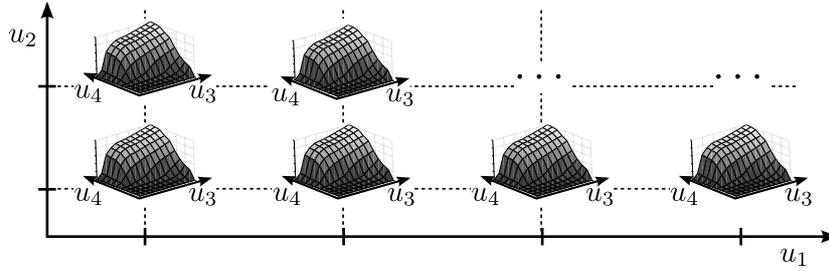
$$\hat{y}_i = \hat{f}_{\text{local},i}(u_3, u_4). \quad (2.26)$$

Hence, these local look-up tables describe local relations (e. g. the dependency of the intake pressure  $p_{2i}$  on the actuator positions of the turbocharger  $s_t$  and the egr valve  $s_{\text{egr}}$  for a fixed engine

operation point,  $p_{2i} = \hat{f}_{\text{local}}(s_t, s_{\text{egr}})$ . Substituting eq. (2.26) in eq. (2.25) gives

$$\hat{y} = \sum_{i=1}^L \Phi_i(u_1, u_2) \hat{f}_{\text{local},i}(u_3, u_4). \quad (2.27)$$

If the number of parameters is regarded, it can be seen that this structure is identical to a four dimensional look-up table. It has therefore the drawback that a relative large amount of parameters is required. Figure 2.6 shows an interpretation of such a four dimensional look-up table.



**Figure 2.6:** Global-local look-up table structure. For each grid point in the global look-up table a further local look-up table is applied. This is especially then suited, if the global model inputs  $u_1$  and  $u_2$  are selected as operation point. This structure is identical to a four dimensional look-up table.

### 2.1.4 Inversion of look-up tables

The inversion of a look-up table is required for some open loop control applications, such as the determination of the required fuel quantity from a requested torque, as shown in Sect. 6.2. The inverse of a look-up table  $f_{\text{base}}(u_1, u_2)$  can be defined for both input variables. Exemplary the inversion with regard to  $u_2$  is considered,

$$u_2 = f_{\text{base}}^{-1}(u_1, y). \quad (2.28)$$

This inverse can be formulated if and only if the look-up table  $f_{\text{base}}(u_1, u_2)$  is strictly monotonic increasing in  $u_2$ ,

$$w_{i,j} < w_{i,j+1} \quad \forall j \in 1 \dots L_2 - 1, \quad (2.29)$$

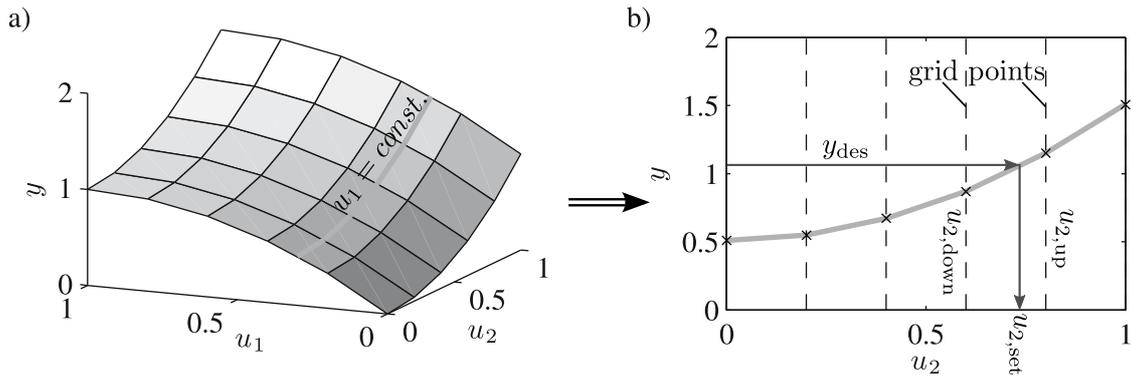
respectively strictly monotonic decreasing,  $w_{i,j} > w_{i,j+1}$ . Hence, if a look-up table needs to be inverted for an application, eq. (2.29) needs to be ensured already at identification of the look-up table. If it is fulfilled, there are in general two approaches to evaluate the inverse of a look-up table.

#### 1. Inverse evaluation of a look-up table

For an inversion with regard to  $u_2$ , values for  $u_1$  and  $y$  need to be given. Applying the value for  $u_1$ , the look-up table can be reduced to a characteristic curve  $y = f_{\text{base},u_1}(u_2)$ , see grey line Fig. 2.7 a). Since the look-up table is strictly monotonic increasing respectively decreasing in  $u_2$ , so is the characteristic curve. Hence, the characteristic curve is invertible such that the inverse can be determined by evaluating  $u_2 = f_{\text{base},u_1}^{-1}(y)$ , as sketched in Fig. 2.7 b).

## 2. Storing the inverse of a look-up table

Alternatively, an inverse model  $u_2 = \hat{f}_{\text{base}}^{-1}(u_1, y)$  can be stored, such that  $\hat{f}_{\text{base}}^{-1} \approx f_{\text{base}}^{-1}$ . To identify such an inverse look-up table, the output  $y$  needs to be discretised. The inverse model can then either be trained with the same training data as applied to train  $f_{\text{base}}(u_1, u_2)$  by swapping the input  $u_2$  and the output  $y$ , or by an inverse evaluation of the look-up table as in approach 1 on the grid points of  $\hat{f}_{\text{base}}^{-1}(u_1, y)$ .



**Figure 2.7:** Inversion of a two-dimensional look-up table with regard to  $u_2$ . a) The two-dimensional look-up table is reduced to a characteristic curve (grey line) with regard to  $u_1 = \text{const.}$  b) Then, the value for  $u_2$  is determined by an inverse evaluation with regard to  $y_{\text{des}}$ . Therefore, the upper and lower grid points,  $u_{2,\text{up}}$  and  $u_{2,\text{down}}$ , need to be determined.

The determination of the inverse with approach 1 provides accurate values, coefficient of determination  $R^2 = 1$ , whereas deviations occur for the second approach,  $\hat{f}_{\text{base}}^{-1} \approx f_{\text{base}}^{-1}$ . The magnitude of the deviation depends on the shape of the look-up table. For the torque model in Fig. 2.2 b) the quality of the inverse trained model is  $R^2 = 0.997$ . This relative small deviation is because of the uniform gradient over the look-up table. If there are considerable variations in the gradient, the deviations are more significant, especially for flat areas, see [75]. Furthermore, if a model is adapted during engine operation, the inverse model needs to be adapted in the same manner. This can deteriorate the quality of the inverse approximation and complicates the storing of an inverse model.

The computational effort is slightly higher for approach 1, since a characteristic curve needs to be determined before the inverse evaluation, see Fig. 2.7. This computational effort can be reduced, if the characteristic curve is only evaluated in the relevant region, but approach 2 is still faster to evaluate. The differences in accuracy and computing time between the two approaches are relatively small. Hence, if the inverse need to be known exactly or if the model is adapted during model operation, approach 1 is recommended and approach 2 can be applied otherwise. A detailed investigation of the computing times of the two approaches can also be found in [145].

## 2.2 Polynomial models and Design of Experiments

Polynomial models are motivated by the Taylor series of the true function  $y = f(\mathbf{u})$ , which is according to [23] given in the neighbourhood of  $\mathbf{u}_0$  by

$$f(\mathbf{u}_0 + \Delta\mathbf{u}) = f(\mathbf{u}_0) + \sum_{i=1}^o \frac{1}{i!} \left( \frac{\partial}{\partial u_1} \Delta u_1 + \frac{\partial}{\partial u_2} \Delta u_2 + \dots + \frac{\partial}{\partial u_p} \Delta u_p \right)^i f(\mathbf{u}_0) + R_o. \quad (2.30)$$

with

$$\begin{aligned} \Delta\mathbf{u} &= \mathbf{u} - \mathbf{u}_0 \\ \mathbf{u}^T &= [u_1, u_2, \dots, u_p] \\ o &\text{ the order of the approximation} \\ R_o &\text{ the remainder.} \end{aligned} \quad (2.31)$$

The higher the order  $o$  the smaller the remainder  $R_o$  and hence the more precise the approximation. However, an increase of the order also increases the number of terms and therefore the complexity of the model.

From eq. (2.30) the general polynomial approximation with  $k$  regressors results

$$\hat{f}(\mathbf{u}) = w_0 + w_1 u_1 + w_2 u_2 + \dots + w_p u_p + w_{p+1} u_1^2 + w_{p+2} u_1 u_2 + \dots + w_k u_p^o. \quad (2.32)$$

The derivatives in eq. (2.30) correspond to the parameters in eq. (2.32), e. g.  $\frac{\partial^i f(\mathbf{u}_0)}{\partial u_j^i}$  corresponds to  $w_j$ , and the deltas in eq. (2.30) correspond to the regressors in eq. (2.32), e. g.  $\Delta u_j^i$  corresponds to  $u_j^i$ . If the true function  $f(\mathbf{u})$  is known, the parameters  $w_j$  can be calculated. However, the true function is in general unknown, why the parameters  $w_j$  in eq. (2.32) are identified by minimising an optimisation criterion with regard to a dataset.

Given a dataset with  $N$  data points, the optimisation criterion can be defined as the sum of squared errors

$$J = \sum_{i=1}^N \left( f(\mathbf{u}(i)) - \hat{f}(\mathbf{u}(i)) \right)^2. \quad (2.33)$$

Since the true values  $f(\mathbf{u}(i))$  are unknown, measured values  $y(i)$  at  $\mathbf{u}(i)$  are employed. As measurements are affected by sensor variations and other uncertainties, an error  $e(i)$  is introduced

$$f(\mathbf{u}(i)) = y(i) + e(i). \quad (2.34)$$

Inserting eq. (2.34) and eq. (2.32) in eq. (2.33) and minimising  $J$  with regard to the parameter vector  $\mathbf{w}$  leads to a linear optimisation problem. Applying the matrix notation with the input matrix  $\mathbf{X}$  consisting of the regressors, the output vector  $\mathbf{y}$  consisting of the measured outputs  $y(i)$  and the

parameter vector  $\mathbf{w}$ ,

$$\mathbf{X} = \begin{pmatrix} 1 & u_1(1) & u_2(1) & \dots & u_p(1) & u_1^2(1) & u_1(1)u_2(1) & \dots & u_p^o(1) \\ 1 & u_1(2) & u_2(2) & \dots & u_p(2) & u_1^2(2) & u_1(2)u_2(2) & \dots & u_p^o(2) \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & u_1(N) & u_2(N) & \dots & u_p(N) & u_1^2(N) & u_1(N)u_2(N) & \dots & u_p^o(N) \end{pmatrix}, \quad (2.35a)$$

$$\mathbf{y} = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{pmatrix} \quad \text{and} \quad \mathbf{w}^T = [w_1 \ w_2 \ \dots \ w_k], \quad (2.35b)$$

the cost function in eq. (2.33) can be written in the form as in eq. (2.7). Hence the parameters  $\mathbf{w}$  can be estimated as in eq. (2.8) by the least squared solution, see e. g. [34, 35, 50, 67, 92],

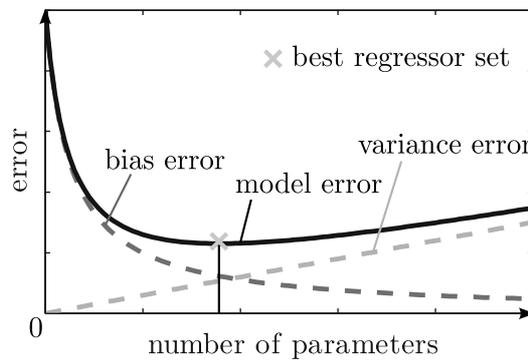
$$\hat{\mathbf{w}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (2.36)$$

This estimate is again a Markov estimate under the assumption that  $\mathbf{e} \sim \mathcal{N}(0, \mathbf{I}\sigma^2)$ . The quality of the estimation depends on the one hand on the noise variance  $\sigma^2$ , since according to [50, 67] the covariance of the estimated parameter vector is

$$\text{Cov}(\hat{\mathbf{w}}) = \text{E}((\hat{\mathbf{w}} - \text{E}(\hat{\mathbf{w}}))(\hat{\mathbf{w}} - \text{E}(\hat{\mathbf{w}}))^T) \quad (2.37a)$$

$$= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}. \quad (2.37b)$$

From this an estimation error results, which is denoted as variance error, and it increases with the number of parameters to be estimated, see e. g. [50, 92]. The bias error on the other hand is introduced by a limited model flexibility and therefore decreases with the number of parameters. This opposite behaviour is known as *bias-variance dilemma* [50, 92, 118, 154] and is illustrated in Fig. 2.8.



**Figure 2.8:** Bias-variance dilemma. The variance error increases linearly with the number of parameters, whereas the bias error decreases to zero. The addition of these two errors gives the model error. The goal of a regressor selection is to find the best performing regressor set with regard to the model error.

The addition of the two errors gives the model error. To find the minimum of the model error, the significant regressors need to be selected from the Taylor approximation. A criterion to rate the quality of models with different numbers of regressors is presented in the following together with an algorithm for a selection of significant regressor.

### 2.2.1 Selection of regressors

To find the best trade-off between bias and variance error, a selection strategy is applied. The selected regressors must be able to approximate the unknown function, while avoiding redundant regressors. Selecting the  $n$  significant regressors of eq. (2.32), the polynomial approximation can be written as

$$\hat{f}(\mathbf{u}) = w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (2.38)$$

with

$$x_i \in \mathcal{M} = \{1, u_1, u_2, \dots, u_p, u_1^2, u_1u_2, \dots, u_p^o\}. \quad (2.39)$$

$\mathcal{M}$  denotes the set of potential regressors derived from the Taylor series, from which the  $n$  significant regressors need to be selected. It has the size  $|\mathcal{M}| = k$ . The number of elements  $k$  depends on the approximation order  $o$  and the dimensionality of the input  $p$ . Since the regressors are motivated by a Taylor series, the summed up order of the interaction terms must be less or equal to the order  $o$ . For example, the summed up order of the interaction term  $u_1^2u_2$  is three, why such an interaction term is only included if  $o \geq 3$ .

The applied size of the order  $o$  depends on the degree of non-linearity of the process. If no a-priori information about the non-linearity of the process is given, odd degrees for  $o$  are often superior to even degrees [50]. This is due to asymptotic boundary effects, since a process often performs in opposite directions for large positive and negative values of an input. For example, a low air mass flow rates implies high soot emissions, while a high air mass flow rates implies low soot emissions. For the combustion models presented in Chap. 3, an order of  $o = 3$  is applied.

#### Criterion of fit

The utilised criterion of fit to select the  $n$  significant regressors is *Mallows'  $C_p$ -statistic*. It is an estimate of the scaled sum of squared errors  $J/\sigma^2$ , see [97], and under the assumption of i.i.d. measurement errors with zero mean a general formulation is given as in [25] by

$$C_p = \frac{\sum_{i=1}^N (y(i) - \hat{y}(i))^2}{\hat{\sigma}^2} - N + 2(n + 1). \quad (2.40)$$

$y(i)$  is the measured output at  $\mathbf{u}(i)$ ,  $\hat{y}(i)$  the modelled output at  $\mathbf{u}(i)$ ,  $N$  the size of the dataset and  $\hat{\sigma}^2$  the estimated noise variance. The criterion consists of a term for the residual sum of squares, which is the first term in eq. (2.40), and a penalty term for the number of applied parameters  $n$ ,

which is presented by the last term. The residual sum of squares decreases with additional model parameters and is therefore accounting for the bias error. The penalty term on the other hand increases with the number of parameters and is accounting for the variance error. The normalisation of the residual sum of squares by the noise variance  $\hat{\sigma}^2$  is applied to relate these errors to each other.

The estimation of the noise variance is given by the sum of squared errors, with the modelled output  $\hat{y}_k(i)$  using all  $k$  regressors, divided by the degrees of freedom,

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N (y(i) - \hat{y}_k(i))^2}{N - k}. \quad (2.41)$$

If the order  $o$  is chosen sufficiently large, the bias error of  $\hat{y}_k(i)$  can be neglected. Then the sum of squared errors composes only a variance error, why, under the assumption of i.i.d. Gaussian measurement errors, eq. (2.41) gives an unbiased estimation of the variance error [35, 50]. Since a degree of freedom is required for this variance estimation, an one is added in eq. (2.40) to the number of model parameters  $n$  [25].

Besides the formulation as in eq. (2.40), there is also an adjusted formulation of Mallows'  $C_p$ -statistic given in [40]

$$\bar{C}_p = \frac{\sum_{i=1}^N (y(i) - \hat{y}(i))^2}{\hat{\sigma}^2} - N + 2n + \frac{2(k - n + 1)}{N - k - 3}. \quad (2.42)$$

In this formulation the variance estimation is not regarded with an additional degree of freedom as in eq. (2.40), but the criterion is adjusted for the case if the number of measurements  $N$  is not much larger than the number of potential regressors  $k$ . Since for the regarded cases in this thesis, the number of measurements  $N$  is considerably larger than the number of potential regressors  $k$ , the definition as in eq. (2.40) can be applied. It is however recommended to use the definition as in eq. (2.42) in other cases.

There are alternative criteria of fit to Mallows'  $C_p$ -statistic, such as the information criteria  $AIC$ ,  $AIC_c$  or  $BIC$  [4, 25, 161]. They do also introduce a regularisation by adding a penalty term, depending on the number of parameters  $n$ , to the sum of squared errors. They are however more difficult to interpret as they describe a relative difference between probability density functions. Other useful criteria are *cross-validation*, *PRESS-statistic* and *ridge-regression* [50, 60, 112]. Cross-validation splits the training dataset into several datasets, leaves then one dataset out for model training and validates the identified model on the left out dataset. This is repeated with each dataset being applied once for validation such that the model error is given by the averaged validation error. The PRESS-statistic is a cross-validation where the left out datasets consist of one point. Although the formulation given in [112] eases the calculation of the PRESS-statistic, these criteria are computationally intensive. The ridge-regression on the other hand suffers from a heuristic choice of a tuning parameter, which needs to be determined a-priori for model identification. Due to the individual drawbacks, Mallows'  $C_p$ -statistic is applied here. The programmed modelling algorithm can however easily be modified, using one of the alternative criteria.

### Selection algorithm

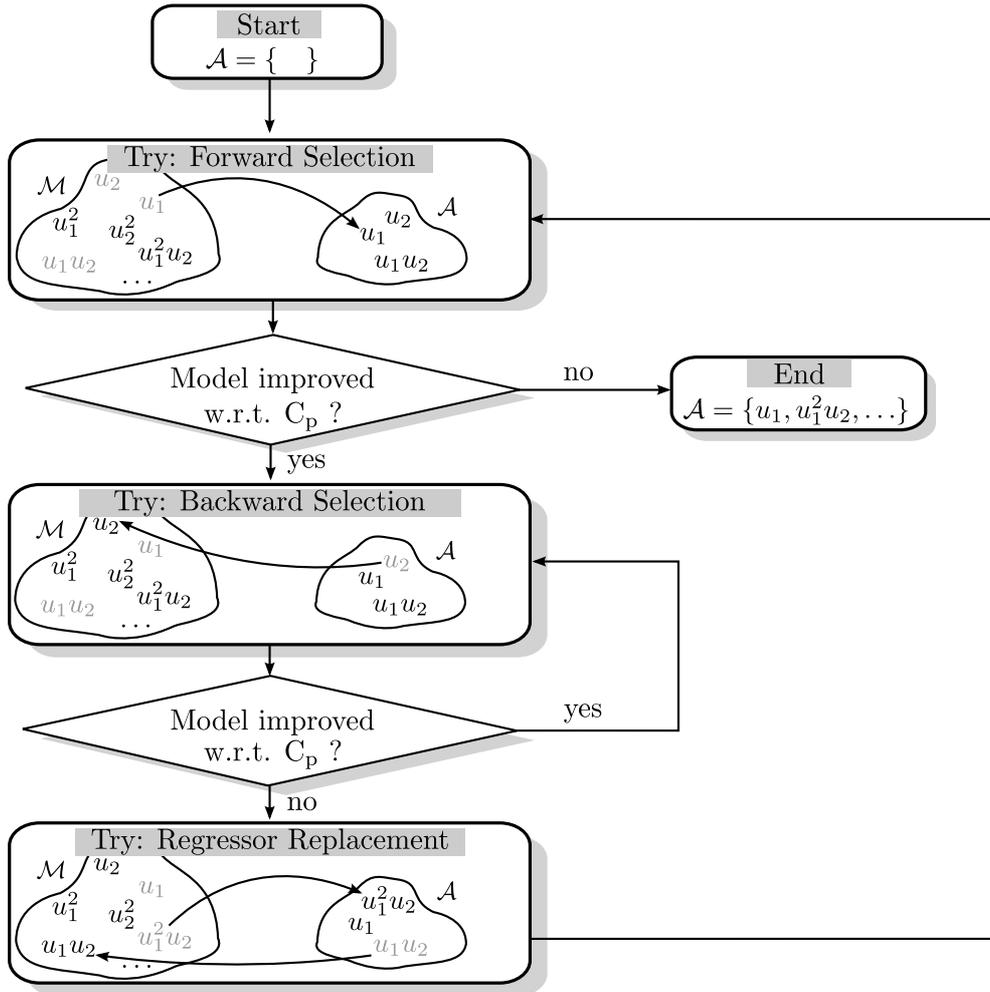
The criterion of fit as in eq. (2.40) enables to compare models with different numbers of regressors. To determine the model with the best regressor set, all possible models need to be identified and compared. This is already for a moderate size of  $\mathcal{M}$  not feasible, why a heuristic selection algorithm is applied. It is a combination of a *forward selection*, *backward selection* and *replacement of regressors* and is depicted in Fig. 2.9. The algorithm starts with the empty set and the forward selection picks the regressors with the highest contribution to the model error. When several parameters are picked, it is possible that an early picked regressor has no longer a significant contribution to the model. Such a regressor is then eliminated by the backward selection. The replacement of regressors enables to exchange a regressor from the set of picked regressors  $\mathcal{A}$  with a regressor from the set of potential regressors  $\mathcal{M}$ , if this has a higher contribution to the model. The individual steps are called several times during the selection algorithm. The frequency and order of the steps are adapted to the number of potential regressors, such that the computationally expensive steps, such as the replacement of regressors, are not called in every selection loop. The individual selection steps are also discussed in more detail in [102]. .

An alternative common used selection algorithm is *orthogonal least squares* [28, 29, 102]. It is fast to calculate, but since the results are identical to a simple forward selection, it is less flexible than the selection algorithm introduced in Fig. 2.9. Another approach applying orthogonal regressors is the *principal component analysis* [68, 71]. There, orthogonal linear combinations of regressors are derived from a singular value decomposition and a forward selection is performed with regard to these linear combinations of regressors, see [198]. The obtained model qualities are comparable to the selection algorithm of Fig. 2.9, but the principal component analysis performs faster due to the orthogonal construction of regressors. A drawback is however that after re-transformation of the linear combinations, all coefficients in eq. (2.32) are in general unequal to zero such that no structural information is given. Hence, if the model quality is of interest rather than the model structure, the principal component analysis can be applied, otherwise the selection algorithm of Fig. 2.9 is recommended.

### 2.2.2 Design of Experiments

The measurement design can be adapted to the applied model structure by the *Design of Experiments (DoE)*. Therefore, a statistical criterion is applied to distribute the measurements in the input space. The most common criterion for polynomial models is the *D-optimal* design, where D stands for determinant. To derive a D-optimal measurement design, a set of potential measurements is assumed  $\mathcal{N}$ . From this, a measurement set with a predefined number of elements is selected  $\mathcal{B} \subseteq \mathcal{N}$  such that the determinant of  $\mathbf{X}^T \mathbf{X}$  is maximised. This task can be written as

$$\max_{\mathcal{B} \subseteq \mathcal{N}} (\det(\mathbf{X}(\mathcal{B})^T \mathbf{X}(\mathcal{B}))). \quad (2.43)$$



**Figure 2.9:** Selection algorithm to determine the best set of regressors  $\mathcal{A}$  from the set of potential regressors  $\mathcal{M}$ . The call of the regressor replacement module is adapted to the size of  $\mathcal{M}$ .

The input matrix  $\mathbf{X}$  depends on the measurement set  $\mathcal{B}$  and also implies information about the applied model structure, which is contained in the columns of  $\mathbf{X}$ , see eq. (2.35). Thus, this criterion is especially suited if a model structure, such as eq. (2.38), already exists.

The maximisation of the determinant  $(\mathbf{X}^T \mathbf{X})$  is according to [78] equivalent to the minimisation of the maximum variance for an evaluation of the model, which is

$$\min_{\mathcal{B} \subseteq \mathcal{N}} \max_{\mathbf{x}} \text{Cov}(\hat{y}) = \min_{\mathcal{B} \subseteq \mathcal{N}} \max_{\mathbf{x}} \text{Cov}(\mathbf{x}^T \hat{\mathbf{w}}) \quad (2.44)$$

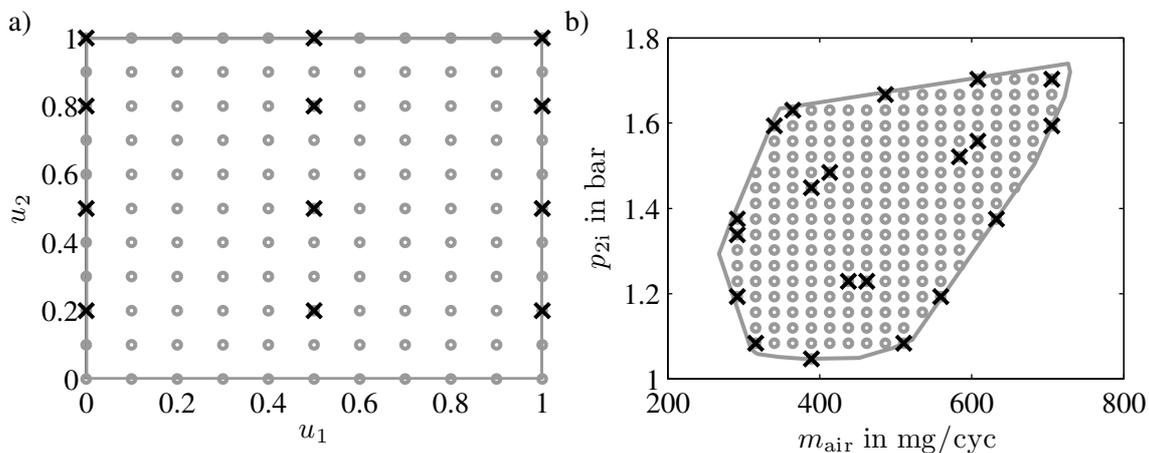
$$= \min_{\mathcal{B} \subseteq \mathcal{N}} \max_{\mathbf{x}} \mathbf{x}^T \text{Cov}(\hat{\mathbf{w}}) \mathbf{x} \quad (2.45)$$

$$= \min_{\mathcal{B} \subseteq \mathcal{N}} \max_{\mathbf{x}} \mathbf{x}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x} \sigma^2, \quad (2.46)$$

where  $\mathbf{x}^T = (x_1, x_2, \dots, x_n)$  is the input vector, see eq. (2.38). Note that for reasons of a clear presentation, the variance of the scalar  $\hat{y}$  is here denoted by its covariance, as for scalars holds that  $\text{Var}(\hat{y}) = \text{Cov}(\hat{y})$ . To guarantee the best measurement set  $\mathcal{B}$  with regard to the applied model

structure, all feasible subsets of  $\mathcal{N}$  need to be compared. As the comparison of all possible subsets is again not feasible, a selection algorithm, such as the one presented in Fig. 2.9, is applied. Since the selection algorithm is heuristic, there is no guarantee that the best measurement set is found. It is therefore recommended to run the selection algorithm several times with a random initialisation and selecting the measurement set with the largest determinant. An overview of selection algorithms for DoE is given in [42].

If a model structure does not exist for the DoE, an assumption about the non-linearity in each dimension has to be made by choosing an individual polynomial order for each dimension. Figure 2.10 a) shows the D-optimal design for an order assumption of  $o = 2$  in  $u_1$  and  $o = 4$  in  $u_2$ , without regarding interaction terms of these inputs. The potential measurements are shown as grey circles and the selected measurements as black crosses. It can be seen that the D-optimal design favours measurements on the boundary of the input space and that more measurements are distributed in the more non-linear assumed input  $u_2$ .



**Figure 2.10:** a) Potential measurements (grey circles) and selected measurements (black crosses) for a D-optimal design with the non-linearity assumption  $o = 2$  in  $u_1$  and  $o = 4$  in  $u_2$ . b) D-optimal design for a local model of the  $\text{NO}_x$  emissions,  $\dot{m}_{\text{nox}}$  at  $n_{\text{eng}} = 2500$  rpm and  $u_{\text{inj}} = 20 \text{ mm}^3/\text{cyc}$ . The local model structure is a polynomial for which the significant regressors are selected. Thus, the measurement design can be adapted to the model structure for an efficient model calibration. Potential measurements (grey circles) are distributed over the drivability space (grey line surrounding the measurements).

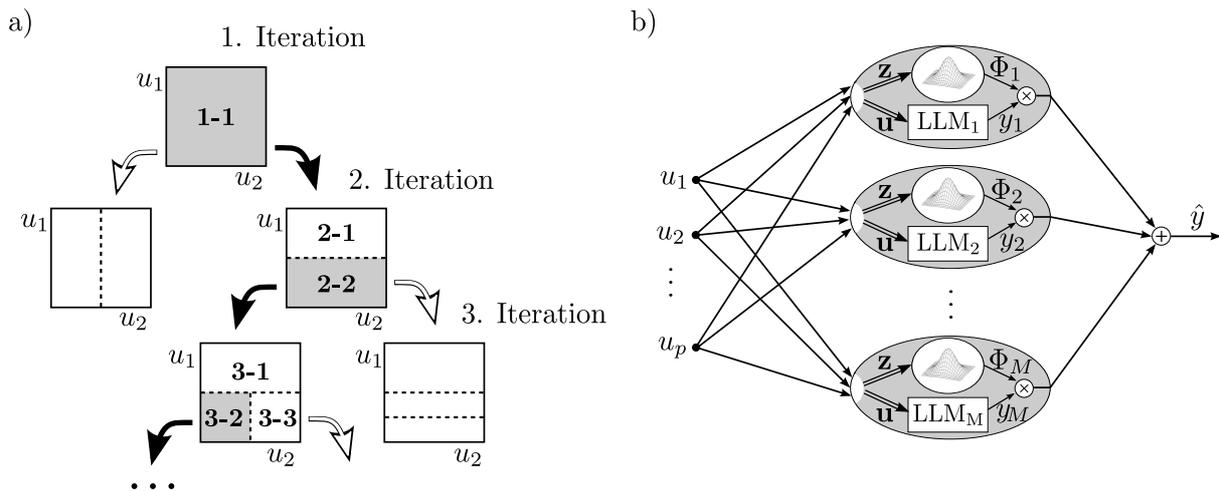
The D-optimal design for an existing model structure is shown in Fig. 2.10 b). A local polynomial model of the  $\text{NO}_x$  emissions is regarded for a fixed engine operation point. The input space of the inputs air mass per cycle  $m_{\text{air}}$  and intake pressure  $p_{2i}$  are shown. The local model is a polynomial of order three for which the significant regressors are selected. The potential measurements are distributed over the drivability space. From these, the measurements are selected such that eq. (2.43) is maximised with regard to the local polynomial model structure. Thus, a problem-oriented distribution of measurements is given, which minimises the variance error of parameter estimation.

It can again be seen that the D-optimal criterion favours measurements on the boundary of the input space. Alternative criteria to the D-optimal design optimise the trace, the largest eigenvalue or the average of  $(\mathbf{X}^T \mathbf{X})$ , see e. g. [42]. The D-optimal criterion is however the only of these that is invariant to transformations of the input variables [130], why it is the preferred criterion for black-box models. A further alternative is the *space filling design*, which optimises the coverage of the input space [142].

The D-optimal distribution of measurements is a well known approach in automotive industry. Because of the increased number of inputs coming from the increased engine complexity, there are long test runs to calibrate an engine. Since test bed time is expensive, the goal is to maximise the information content of the measurement data. This is why D-optimal designs have been established in the last decade and are meanwhile implemented in commercial measurement tools [22, 86, 104, 137, 158, 162]. Besides this, there are also techniques to distribute measurements for multi-model approaches, see [100, 189].

## 2.3 LOLIMOT

LOLIMOT is an abbreviation for *local linear model tree* [123], which is developed for a MISO-model. Its local models are of linear or, if an offset is applied, affine type and they are weighted to an overall model output by means of a Gaussian weighting function. Thus, a smooth transition from one local model to another is given. The partition of the input space, where the local models are valid, is determined by a tree construction algorithm applying axis-orthogonal splits, see Fig. 2.11 a). The global model output is then calculated by the sum of the local model output multiplied by the weighting function, see Fig. 2.11 b).



**Figure 2.11:** a) Tree construction algorithm and b) model structure of LOLIMOT. The local models (LLM) are of linear or affine type and are superimposed to a global model by means of Gaussian weighting functions. The partitions where the local models are valid are determined by the tree construction algorithm applying axis-orthogonal splits.

The weighting function depends on the  $z$ -regressor, which is in general a subset of the input space,  $\mathbf{z} \subseteq \{u_1, u_2, \dots, u_p\}$ . The global model output can then be written as

$$\hat{y}_{\text{LOLIMOT}}(\mathbf{u}) = \sum_{j=1}^M \Phi_j(\mathbf{z}) \hat{y}_j(\mathbf{u}) \quad (2.47a)$$

$$\text{with } \hat{y}_j(\mathbf{u}) = (w_{0,j} + w_{1,j}u_1 + \dots + w_{p,j}u_p), \quad (2.47b)$$

where  $M$  is the number of local linear models,  $\mathbf{w}_j^T = [w_{0,j}, w_{1,j}, \dots, w_{p,j}]$  are the  $j$ -th local model parameters and  $\Phi_j$  is the normalised Gaussian weighting function of the  $j$ -th local model,

$$\Phi_j(\mathbf{z}) = \frac{\exp\left(-\frac{1}{2} \left\| \frac{\mathbf{z} - \mathbf{z}_{0,j}}{\sigma_j^2} \right\|_2^2\right)}{\sum_{j=1}^M \exp\left(-\frac{1}{2} \left\| \frac{\mathbf{z} - \mathbf{z}_{0,j}}{\sigma_j^2} \right\|_2^2\right)}. \quad (2.48)$$

$\mathbf{z}_{0,j}$  is the centre coordinate of the  $j$ -th local model and  $\sigma_j^2$  is the variance of the Gaussian. Both are determined by the tree construction algorithm, shown in Fig. 2.11 a). The centres are each placed in the middle of the local models and the variances of the Gaussian are determined by a vector depending on the width of a local model in the respective dimensions. The parameters for the  $j$ -th local model are determined by minimizing the weighted sum of squared errors

$$J_j = \mathbf{e}^T \mathbf{Q}_j \mathbf{e}, \quad \text{with } \mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}_j. \quad (2.49)$$

Writing eq. (2.47) with the applied  $N$  training values in matrix notation and minimising the sum of squared errors in eq. (2.49) leads to the local parameter estimation for the  $j$ -th local model, see e. g. [67, 118],

$$\hat{\mathbf{w}}_j = (\mathbf{X}_j^T \mathbf{Q}_j \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{Q}_j \mathbf{y}. \quad (2.50)$$

The weighting matrix of the  $j$ -th local model  $\mathbf{Q}_j$  is a diagonal matrix with elements  $\Phi_j(\mathbf{z})$ ,

$$\mathbf{Q}_j = \begin{pmatrix} \Phi_j(\mathbf{z}(1)) & 0 & \dots & 0 \\ 0 & \Phi_j(\mathbf{z}(2)) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Phi_j(\mathbf{z}(N)) \end{pmatrix}. \quad (2.51)$$

The input matrix of the  $j$ -th local model  $\mathbf{X}_j$  contains the  $p$  inputs as columns and a column with one's for the offset if affine local models are applied, compare to eq. (2.47), and the output vector  $\mathbf{y}$  the measured output values,

$$\mathbf{X}_j = \begin{pmatrix} 1 & u_1(1) & u_2(1) & \dots & u_p(1) \\ 1 & u_1(2) & u_2(2) & \dots & u_p(2) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & u_1(N) & u_2(N) & \dots & u_p(N) \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{pmatrix}. \quad (2.52)$$

With this notation, the model as in eq. (2.47) can be written as

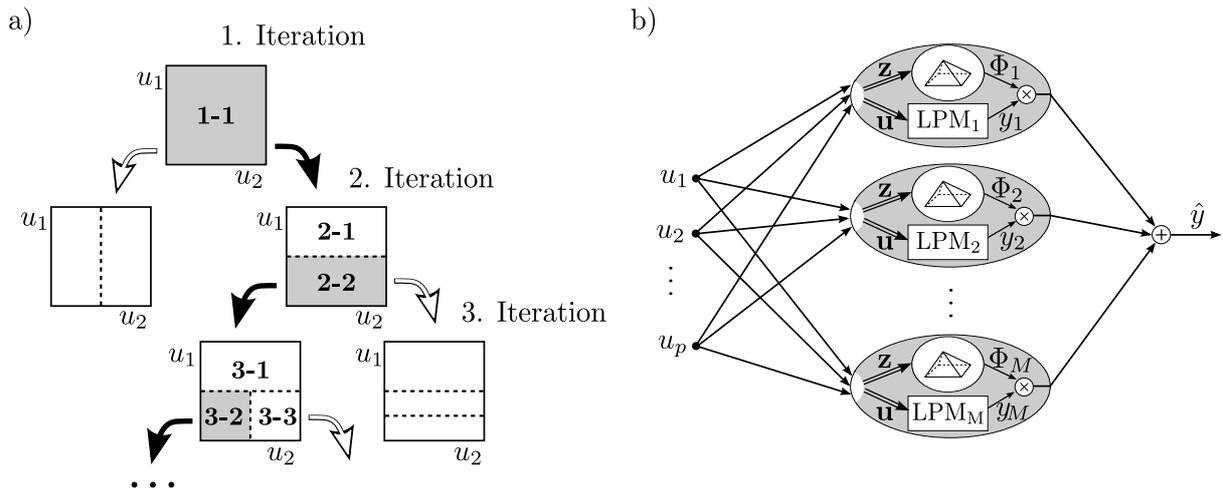
$$\hat{y}_{\text{LOLIMOT}}(\mathbf{u}) = \sum_{j=1}^M \Phi_j(\mathbf{z}) \hat{y}_j(\mathbf{u}) \quad (2.53a)$$

$$= \sum_{j=1}^M \mathbf{Q}_j \mathbf{X}_j \hat{\mathbf{w}}_j. \quad (2.53b)$$

More details about the LOLIMOT algorithm are given among others in [65, 117, 118]. Enhancements to LOLIMOT regarding the identification of dynamic systems can be found in [187]. Similar algorithms to LOLIMOT with various differences in the splitting algorithm and the local parameter estimation are given in [11, 49, 69, 70]. A comparison of a clustering and an axis-orthogonal splitting algorithm is presented in [121].

## 2.4 LOPOMOT

The *local polynomial model tree* (LOPOMOT) is a combination of the presented LOLIMOT structure and the adaptive polynomial models presented before for a MISO model. Like for LOLIMOT, an axis orthogonal splitting algorithm is applied, see Fig. 2.12 a). In contrast to LOLIMOT, local polynomial models are used as local models. For these local polynomials, a local formulation of Mallows'  $C_p$ -statistic is derived to select the significant regressors. This enables an adaption of the local models to the non-linearities of the process and reduces the model error. Thus, the positive properties of LOLIMOT and the adaptive polynomials are combined.



**Figure 2.12:** LOPOMOT with a) the axis-orthogonal splitting algorithm as in the LOLIMOT algorithm and b) the model structure consisting of adaptive local polynomial models (LPM) that are weighted by means of a pyramidal weighting function to a global model output. The partitions where the local models are valid are determined by the tree construction algorithm.

A related approach is presented in [12]. There the determination of the local polynomial order is included in the splitting algorithm. The improvement of increasing the local polynomial order is compared to a model split and the superior is performed. In contrast to the following presentation, the employed selection criteria are based on heuristic choices of tuning parameters.

In the following, the LOPOMOT structure is introduced. Thereafter, a local formulation of Mallows'  $C_p$ -statistic is derived for the local regressor selection algorithm. Based on that, a selection criterion for the best model partition is presented. Furthermore, the identification of dynamic systems is discussed. A comparison to LOLIMOT using a stationary and a dynamical example shows the advantages of LOPOMOT.

Similar to eq. (2.47) the local model structure of LOPOMOT can be written as

$$\hat{y}_{\text{LOPOMOT}}(\mathbf{u}) = \sum_{j=1}^M \Phi_j(\mathbf{z}) \hat{y}_j(\mathbf{u}) \quad (2.54a)$$

$$\text{with } \hat{y}_j(\mathbf{u}) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n \quad (2.54b)$$

$$\text{and } x_i \in \mathcal{M} = \{1, u_1, u_2, \dots, u_p, u_1^2, u_1 u_2, \dots, u_p^o\}, \quad (2.54c)$$

as in eq. (2.38). The weighting function  $\Phi_j(\mathbf{z})$  can be a Gaussian as for the LOLIMOT model, eq. (2.48), or any other function that decays to zero. For the emission models presented in Chap. 3, a pyramidal weighting function is applied, see App. D.3. In contrast to the Gaussian weighting function, it has compact support and is equivalent to a linear interpolation for an equidistant lattice structure. This eases the model simulation, especially for an implementation on an ECU.

The structure of LOPOMOT is depicted in Fig. 2.12. A comparison to LOLIMOT in Fig. 2.11 shows that the tree construction algorithms are the same. The major differences are given by the determination of the local models. LOPOMOT applies adaptive local polynomial models (LPM), whereas LOLIMOT applies local linear models (LLM). There are further differences in the identification algorithm, such as the selection of regressors or the selection of the best model partition, which are presented in the following.

### 2.4.1 Local selection of regressors

A criterion is required for the local regressor selection that compares the local model qualities and takes their complexities into account. Such a criterion is given by the local formulation of Mallows'  $C_p$ -statistic. To rate the local model complexities, the effective number of parameters [106] are calculated for the local linear regression model as shown e. g. in eq. (2.53) by [96]

$$n_{\text{eff},j} = \text{trace}(\mathbf{H}_j) = \quad (2.55a)$$

$$= \text{trace} \left( \mathbf{Q}_j \mathbf{X}_j (\mathbf{X}_j^T \mathbf{Q}_j \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{Q}_j \right) \quad (2.55b)$$

$$= \text{trace} \left( (\mathbf{X}_j^T \mathbf{Q}_j \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{Q}_j^2 \mathbf{X}_j \right), \quad (2.55c)$$

with

$$\mathbf{H}_j = \mathbf{Q}_j \mathbf{X}_j (\mathbf{X}_j^T \mathbf{Q}_j \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{Q}_j. \quad (2.56)$$

$\mathbf{H}$  is called the *hat* or projection matrix as  $\hat{\mathbf{y}}_j = \mathbf{H}_j \mathbf{y}$ , since with the model definition as in eq. (2.53) and the parameter estimation as in eq. (2.50) follows

$$\mathbf{H}_j \mathbf{y} = \mathbf{Q}_j \mathbf{X}_j (\mathbf{X}_j^T \mathbf{Q}_j \mathbf{X}_j)^{-1} \mathbf{X}_j^T \mathbf{Q}_j \mathbf{y} \quad (2.57a)$$

$$= \mathbf{Q}_j \mathbf{X}_j \hat{\mathbf{w}}_j \quad (2.57b)$$

$$= \hat{\mathbf{y}}_j. \quad (2.57c)$$

The weighting matrix of the  $j$ -th local model  $\mathbf{Q}_j$  ( $N \times N$ ) is a diagonal matrix with elements  $\Phi_j(\mathbf{z})$  evaluated at the  $N$  data points, see eq. (2.51). The input matrix of the  $j$ -th local model  $\mathbf{X}_j$  ( $N \times n$ ) contains the  $n$  applied regressors, see eq. (2.54), evaluated at the  $N$  data points,

$$\mathbf{X}_j = \begin{pmatrix} x_1(1) & x_2(1) & \dots & x_n(1) \\ x_1(2) & x_2(2) & \dots & x_n(2) \\ \vdots & \vdots & \dots & \vdots \\ x_1(N) & x_2(N) & \dots & x_n(N) \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{pmatrix}. \quad (2.58)$$

The reformulation of the trace in eq. (2.55c) follows from  $\text{trace}(AB) = \text{trace}(BA)$  and allows a less expensive computation of the effective number of parameters. Note that the effective number of parameters simplifies for  $\mathbf{Q}_j = \mathbf{I}$  to  $\text{trace}(\mathbf{H}) = p$ , where  $p$  is the number of parameters and the dimension of the projection space of the projection matrix  $\mathbf{H}$  [50]. Given this rating of model complexity and the general formulation of Mallows'  $C_p$ -statistic in eq. (2.40), the local formulation can be stated similar to [96] as

$$C_{p,\text{local},j} = \frac{(\mathbf{y} - \hat{\mathbf{y}}_j)^T \mathbf{Q}_j (\mathbf{y} - \hat{\mathbf{y}}_j)}{\hat{\sigma}_j^2} + \text{trace}(\mathbf{Q}_j) + 2(n_{\text{eff},j} + 1), \quad (2.59)$$

where  $\hat{\mathbf{y}}_j$  is the local model output,  $\hat{\sigma}_j^2$  the local noise variance and  $\text{trace}(\mathbf{Q}_j)$  can be regarded as the local effective number of measurements. Analogue to eq. (2.41), the local variance is determined by the local model with all  $k$  potential regressors  $\hat{\mathbf{y}}_{j,k}$ ,

$$\hat{\sigma}_j^2 = \frac{(\mathbf{y} - \hat{\mathbf{y}}_{j,k})^T \mathbf{Q}_j (\mathbf{y} - \hat{\mathbf{y}}_{j,k})}{\text{trace}(\mathbf{Q}_j) - n_{\text{eff},j,k}}. \quad (2.60)$$

Here the local degrees of freedom are determined by the difference between the effective number of measurements  $\text{trace}(\mathbf{Q}_j)$  and the effective number of parameters  $n_{\text{eff},j,k}$ . Again, a one is added to the effective number of parameters  $n_{\text{eff},j}$  in eq. (2.59) to account for this variance estimation. Since the added one is a constant in eq. (2.59), it has no influence on the local regressor selection. It is however important for the global model selection, as is shown in the following section. Note that it is possible to force more or less regressors in the local models, by decreasing or increasing the coefficient of  $n_{\text{eff},j}$  in eq. (2.59). However, the choice of 2 is recommended from a theoretical point of view.

### 2.4.2 Global selection of model partition

Based on the tree construction algorithm of Fig. 2.12 a), several model partitions are calculated until a predefined termination condition is reached. To select the model with the best number of local models, again a formulation of Mallows'  $C_p$ -statistic is applied.

Given the local noise variance  $\hat{\sigma}_j^2$  for each local model, estimated by eq. (2.60), an individual noise variance for each measurement  $y(i)$  can be determined by utilising the local weights  $\Phi_j(\mathbf{z}(i))$ ,

$$\hat{\sigma}^2(\mathbf{z}(i)) = \sum_{j=1}^{M_{\max}} \Phi_j(\mathbf{z}(i)) \hat{\sigma}_j^2. \quad (2.61)$$

Since a good variance estimation  $\hat{\sigma}^2(\mathbf{z}(i))$  is crucial for an appropriate model selection, the variance estimation of eq. (2.61) is calculated by the model with the lowest bias error, which is usually the model with the maximum number of partitions  $M_{\max}$ . Using the individual noise variance estimation of eq. (2.61) and the local formulation of the  $C_p$ -statistic, eq. (2.59), the global formulation of the  $C_p$ -statistic results

$$C_{p,\text{global}} = \sum_{i=1}^N \frac{(y(i) - \hat{y}(i))^2}{\hat{\sigma}^2(\mathbf{z}(i))} - N + 2 \sum_{j=1}^M (n_{\text{eff},j} + 1). \quad (2.62)$$

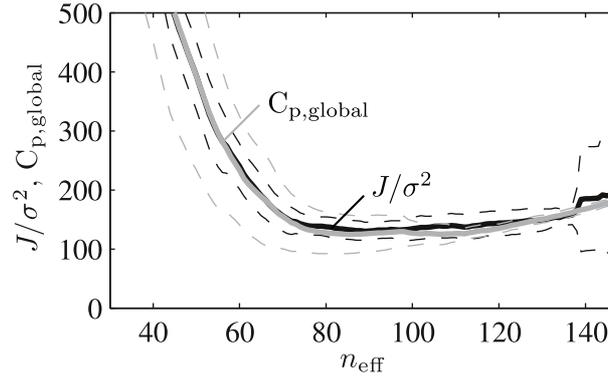
This criterion rates the quality of a model with  $M$  model partitions for  $N$  measured scalar outputs  $y(i)$ . The quality mainly depends on the squared sum of residuals (first term) and the overall number of effective parameters (last term). The addition of an one to the local effective number of parameters in the last term results from the  $M$  local variance estimations. This has an important effect on the global model selection, as thus the number of local models  $M$  is additionally penalised.

The performance of the estimate  $C_{p,\text{global}}$  is illustrated in Fig. 2.13 for data from the example presented in Sect. 2.4.4. Since the true function is known for this example, the scaled sum of squared errors  $J/\sigma^2$  can be calculated using the true output  $f(\mathbf{u}(i))$ , see eq. (2.33). It can be seen that  $C_{p,\text{global}}$  is able to estimate  $J/\sigma^2$  and that it is especially suited to determine the minimum of  $J/\sigma^2$  with regard to the effective number of parameters. This minimum corresponds to the best trade-off between bias and variance error, as sketched in Fig. 2.8. The estimate  $C_{p,\text{global}}$  is therefore well suited to select the best model partition. The dashed lines in the plot indicate the variation in the calculation of  $J/\sigma^2$  and the estimation of  $C_{p,\text{global}}$  for ten Monte-Carlo simulations.

### 2.4.3 Dynamic models

The LOPOMOT algorithm can also be used to model discrete time dynamic processes. Therefore, the basic LOPOMOT net model is taken, but with shifted inputs and outputs

$$y(k) = f(u_1(k), u_1(k-1), \dots, u_1(k-m_{u1}), u_2(k), u_2(k-1), \dots, u_p(k-m_{up}), y(k-1), \dots, y(k-m_y)). \quad (2.63)$$



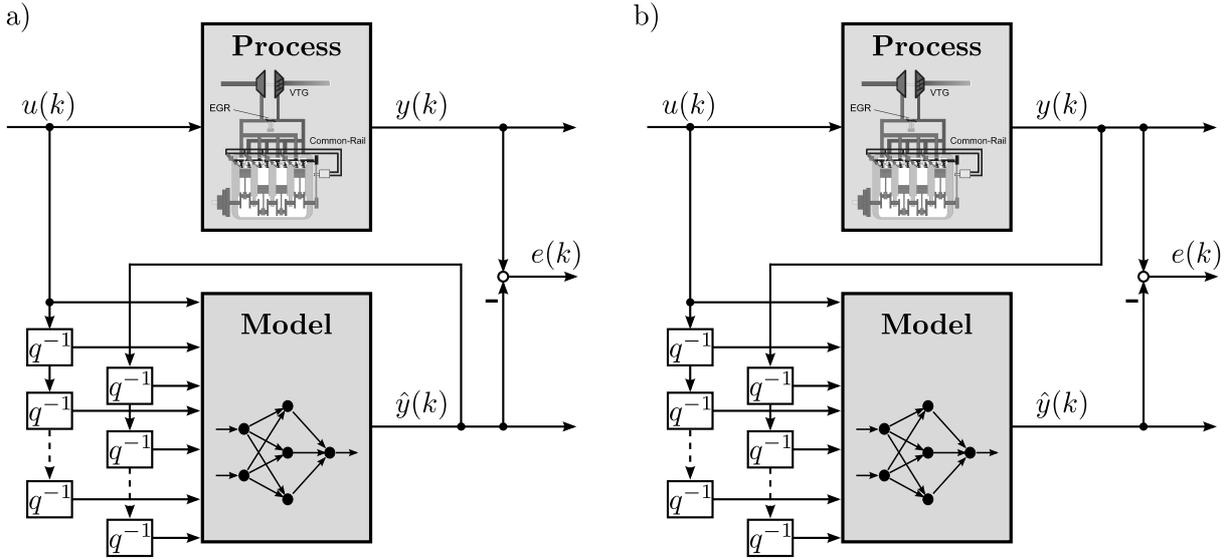
**Figure 2.13:** Scaled sum of squared errors  $J/\sigma^2$ , calculated by eq. (2.33), and its estimate  $C_{p,\text{global}}$ , calculated by eq. (2.62), over the effective number of parameters  $n_{\text{eff}}$ .  $J/\sigma^2$  is calculated from undisturbed validation data and  $C_{p,\text{global}}$  is calculated from the training data, both taken from the stationary example in Sect. 2.4.4. Dashed lines show the variances for ten Monte Carlo simulations.

The feedback of inputs and outputs to a static model structure is known as external dynamics and discussed in detail in [65, 114]. Given eq. (2.63), the regressors for a dynamical model can again be motivated by a Taylor series. The higher the dynamical orders  $m_{u_1}, m_{u_2}, \dots, m_{u_p}$  and  $m_y$  the larger the set of potential regressors  $\mathcal{M}$ . It is therefore recommended to cancel regressors which are assumed to have no influence on the process or are poor to interpret. For example, inputs depending on different time shifts, such as  $u_1(k-1)u_1(k-2)$ , are cancelled for the following example. This eases the interpretation of the regressors and reduces the number of potential regressors.

The parameters can be identified either in parallel or series-parallel model structure [114, 118]. The former uses the modelled outputs as feedback, e. g.  $\hat{y}(k-1)$ , and the latter the measured outputs, e. g.  $y(k-1)$ , see Fig. 2.14. The parallel model structure leads to a non-linear optimisation problem, whereas the series-parallel model structure can be solved by a direct least squares algorithm. [188] shows that the regularisation effect and the interpretation of the local models are lost for an identification with the parallel model structure, why an explicit regularisation is presented there for this case.

Because of the high number of potential regressors, the non-linear optimisation and the explicit regularisation, the parallel model structure is computational intensive, why it is only limited practical for a selection algorithm as is applied for LOPOMOT. Hence, the series-parallel model structure is utilised. Detailed discussions about the selection of regressors for polynomial models with external dynamics in parallel and series-parallel model structure can also be found in [36, 127, 128].

Since the series-parallel model structure optimises the one step ahead prediction, whereas the model usually operates as simulation model, the determination of the sampling time  $T_0$  is crucial for a reliable model operation. It is in [67] recommended to choose the sampling time  $T_0$  subject to  $T_{95}$ ,



**Figure 2.14:** Model structures with external dynamics for a) a parallel model structure and b) a series-parallel model structure. The parallel model structure uses time delayed modelled outputs (e. g.  $\hat{y}(k - 1)$ ) and the series-parallel model structure uses measured outputs (e. g.  $y(k - 1)$ ) as model inputs.

the 95 % settling time of the step response,

$$T_0 = \left\{ \frac{1}{5}, \dots, \frac{1}{15} \right\} T_{95}. \quad (2.64)$$

A too small sampling time is problematic as it yields to a selection of  $y(k - 1)$ , which estimates  $y(k)$  satisfactorily well for small sampling times such that no further parameters can be estimated reliably [2, 128, 198]. On the other hand, a model with a too large sampling time can not describe the dynamic process characteristics properly. Choosing the sampling time subject to eq. (2.64) yields however good results, as can be seen in the following dynamic example. A similar rule of thumb for the determination of  $T_0$  can also be found in [95].

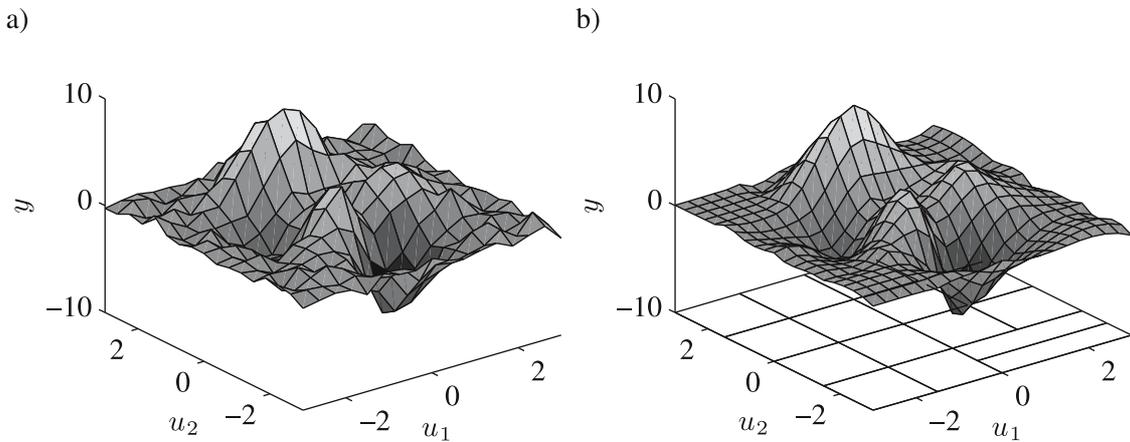
It shall be mentioned that the discussed computational complexity only regards the identification of the model. The simulation on the other hand is relatively fast, since it is restricted to evaluations of polynomials and linear interpolations. Furthermore, the simulation of the LOPOMOT model is also programmed as *c-function* such that it can easily implemented on engine test beds with Rapid Control Prototyping systems or in engine control units. Applying the corresponding *simulink block*, the LOPOMOT model can also be integrated into block-oriented programming, see App. D.4.

#### 2.4.4 Comparison to LOLIMOT

In the following, two examples are presented to illustrate the performance of LOPOMOT in comparison to LOLIMOT. At first, a two dimensional stationary non-linear process is shown, then a dynamical example applying a Hammerstein process is regarded.

### Stationary non-linear process

For creation of measurement data, the MATLAB test function *peaks* is utilised. This function describes a non-linear relation with regard to two input variables. It consists of several Gaussian functions and is in the following identified by a black-box model. Ten Monte-Carlo simulations are given for the test function with Gaussian noise  $\frac{\sigma_{\text{noise}}}{\sigma_y} = 0.278$  added to the measurements. For model training 289 noisy data points and for model validation 2112 undisturbed data points are applied. Figure 2.15 a) shows one realisation of the noisy training data and Fig. 2.15 b) the identified model for this realisation with the model partitions indicated in the  $u_1, u_2$ -plane. The approximation of such a non-linear relation can also be regarded as a parametric approximation of a look-up table.



**Figure 2.15:** a) Training data with noise  $\frac{\sigma_{\text{noise}}}{\sigma_y} = 0.278$  and b) identified model with 16 local polynomial models. The coefficients of determination are  $R^2_{\text{train}} = 0.951$  and  $R^2_{\text{valid}} = 0.971$ , for training respectively validation data. The approximation can also be regarded as parametric approximation of a look-up table. For the local models, regressors up to third order, e. g.  $u_1, u_1^3, u_1 u_2^2, \dots$ , are applied.

It can be seen that the model smooths the noisy measurements and that it is able to simulate the process behaviour. This is also indicated by the coefficients of determination  $R^2$  given in the caption. A comparison to LOLIMOT is shown in Tab. 2.1 for the averaged values of ten Monte Carlo simulations. Both model structures are able to approximate the non-linear process, see  $R^2_{\text{train}}$ . LOLIMOT needs however significantly more partitions (# LM) to achieve similar results to LOPOMOT, which also increases the number of parameters  $n_{\text{total}}$  and  $n_{\text{eff}}$ . The validation results of LOLIMOT are slightly worse than for LOPOMOT. However, since no redundant regressors which do not contribute to the model quality are employed in this example, the differences in model quality are relatively small. Redundant regressors can deteriorate the model quality as is shown in the following dynamic example.

**Table 2.1:** Comparison of LOLIMOT and LOPOMOT for the test function as in Fig. 2.15 with regard to model quality on training data  $R^2_{\text{train}}$ , model quality on validation data  $R^2_{\text{valid}}$ , total number of parameters  $n_{\text{total}}$ , effective number of parameters  $n_{\text{eff}}$  and number of local models # LM. Values are averaged over ten Monte Carlo simulations with Gaussian noise  $\frac{\sigma_{\text{noise}}}{\sigma_y} = 0.278$  added to the training data.

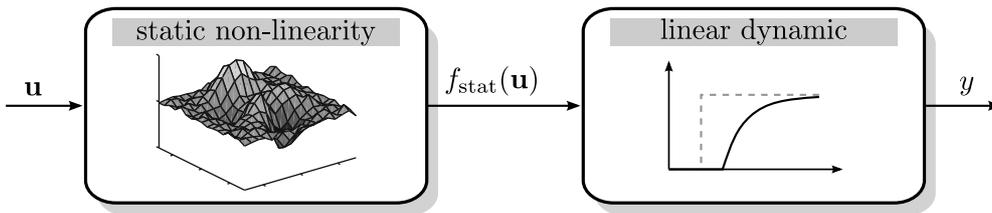
	$R^2_{\text{train}}$	$R^2_{\text{valid}}$	$n_{\text{total}}$	$n_{\text{eff}}$	# LM
LOPOMOT	0.957	0.966	135.6	97.8	21.4
LOLIMOT	0.958	0.958	210.0	125.0	70.0

### Hammerstein process

The dynamic performance is in the following regarded for a Hammerstein process with dead time. The stationary non-linearity is the same as in the previous example and the dynamic transfer function is the third test process taken from [63],

$$G_{\text{dyn}}(s) = \frac{y(s)}{f_{\text{stat}}(u_1, u_2)(s)} = \frac{1 + 2s}{(1 + 10s)(1 + 7s)(1 + 3s)} e^{-4s}. \quad (2.65)$$

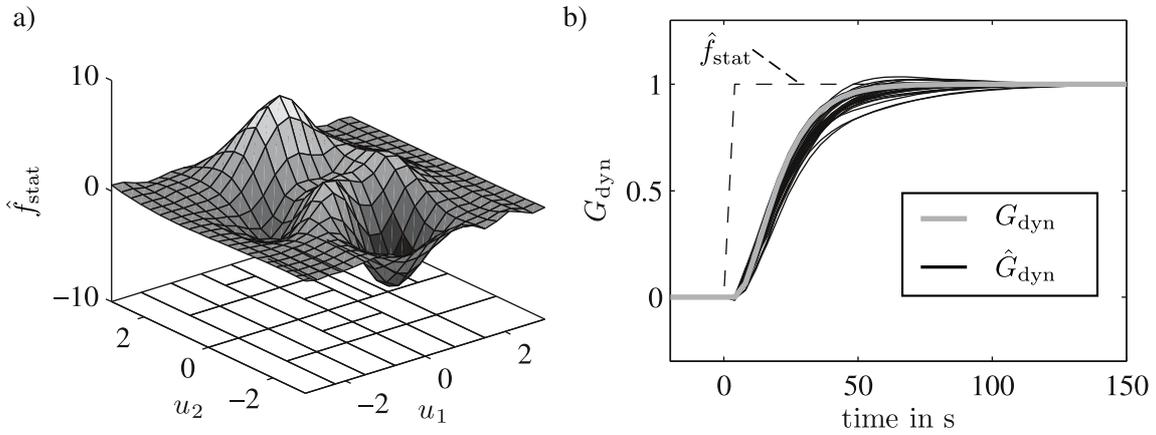
The block diagram of this process is shown in Fig. 2.16. It is again identified by a black-box model utilising the process inputs  $\mathbf{u}$  and the process output  $y$  as measurements. According to eq. (2.64) and with regard to a simulated step response, the sampling time is chosen to  $T_0 = 4.0$  s. For excitation of the process, an amplitude-modulated pseudo-random binary signal (APRBS) [67, 118] is applied on the inputs with the length  $t = 4096$  s. Measurements are again disturbed by noise,  $\frac{\sigma_{\text{noise}}}{\sigma_y} = 0.068$ , and ten Monte Carlo simulations are performed to average the results. The dynamic orders for identification are chosen relatively high to  $m_y = 8$ , and  $m_{u1} = m_{u2} = 8$ , since the information that the system possesses a dead time is not applied for the selection of the orders.



**Figure 2.16:** Block diagram of the Hammerstein process. The non-linearity is the same as in the stationary example and the linear dynamic is given in eq. (2.65) with a dead time. The Hammerstein process is identified as a black-box model with measurements for the input  $\mathbf{u}$  and the output  $y$ .

The identification results for one realisation is shown in Fig. 2.17. The presented model consists of 30 local models and  $n_{\text{eff}} = 452.8$  effective parameters. The identified stationary gain is presented in Fig. 2.17 a) and the identified dynamics in Fig. 2.17 b). The dynamics are normalised to one and

the true unified step response is shown for a comparison in grey. The dead time  $T_d = 4$  s can well be seen in the step responses and the stationary and dynamic behaviour is well estimated.



**Figure 2.17:** a) Identified stationary gain  $\hat{f}_{\text{stat}}$  and b) true linear dynamic  $G_{\text{dyn}}$  (grey) and identified linear dynamics  $\hat{G}_{\text{dyn}}$  (black) for 30 unified step responses in all local models. All steps, also these with a negative gain, are normed to one for a better comparison of the identified dynamics.

The averaged results of ten Monte Carlo simulations are summarised in Tab. 2.2. The local polynomial model structure LOPOMOT is shown with and without applying the selection algorithm. If the selection algorithm is not applied, unstable local models result for some realisations. This is mainly due to a lack of excitation for the redundant regressors. The LOLIMOT model also performs significantly worse than LOPOMOT, while possessing a major number of local models. The LOLIMOT model suffers from the relatively high chosen dynamic model order and therefore from the consideration of redundant regressors. Because of the redundant regressors, there are also unstable local models for LOLIMOT. In contrast, the LOPOMOT model cancels the redundant regressors and performs well on training and validation data.

**Table 2.2:** Comparison of LOLIMOT and LOPOMOT for a Hammerstein test function with regard to model quality on training data  $R_{\text{train}}^2$ , model quality on validation data  $R_{\text{valid}}^2$ , total number of parameters  $n_{\text{total}}$ , effective number of parameters  $n_{\text{eff}}$  and number of local models # LM. Values are averaged over ten Monte Carlo simulations with Gaussian noise  $\frac{\sigma_{\text{noise}}}{\sigma_y} = 0.068$  added to the measurements.

	$R_{\text{train}}^2$	$R_{\text{valid}}^2$	$n_{\text{total}}$	$n_{\text{eff}}$	# LM
LOPOMOT	0.978	0.945	749.0	459.2	29.9
LOPOMOT with all Regs	0.988	0.522	4837.1	2114.7	29.9
LOLIMOT	0.954	0.897	1533.6	949.4	56.8

The LOPOMOT model applies more model partitions than in the stationary example. This is because more number of measurements and also more accurate measurements ( $\frac{\sigma_{\text{noise}}}{\sigma_y} = 0.068$ ) are provided for the second example. A further comparison of LOPOMOT and LOLIMOT for a Hammerstein process with an oscillating  $\text{PT}_2$  can be found in [201].

## 2.5 Kernel models

Kernel models primarily consist of the training data and only a small number of parameters need to be identified. To simulate a model, the measured outputs of the training dataset are weighted by a normalised weighting function and summed up to a model output. Because of the calculation of the weights, Kernel models require a significant computing time for model evaluation. The output of a Kernel model can in general be written by the *Nadaraya-Watson kernel-weighted average*, see [50, 44],

$$\hat{f}_{\text{kernel}}(\mathbf{u}) = \frac{\sum_{i=1}^N K_{\mathbf{h}}(\mathbf{u}, \mathbf{u}(i))y(i)}{\sum_{i=1}^N K_{\mathbf{h}}(\mathbf{u}, \mathbf{u}(i))}, \quad (2.66)$$

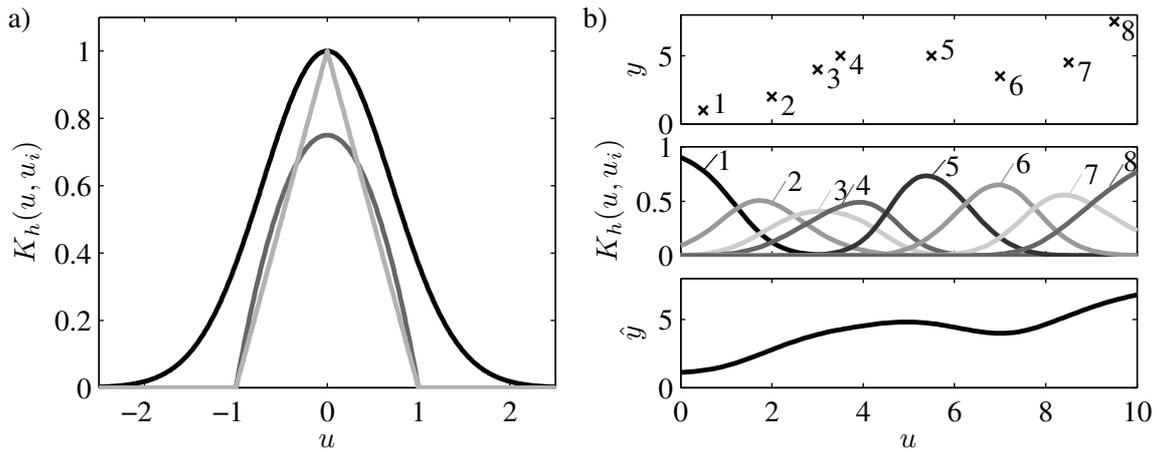
where  $\mathbf{u}$  is the model input,  $(\mathbf{u}(i), y(i))$  are the  $N$  stored data points from model training and  $K_{\mathbf{h}}(\cdot)$  is the kernel function, here given as the Gaussian

$$K_{\mathbf{h}}(\mathbf{u}, \mathbf{u}(i)) = \exp\left(-\frac{1}{2} \left\| \frac{\mathbf{u} - \mathbf{u}(i)}{\mathbf{h}^2} \right\|_2^2\right). \quad (2.67)$$

Such Kernel models are also called *Radial Basis Functions*. The kernel function in eq. (2.67) depends on the bandwidth parameter  $\mathbf{h}$ , which is the only parameter to be identified. The bandwidth is here determined by minimising the cross-validation error on the training data using non-linear optimisation techniques. Other kernel functions, such as the *Epanechnikov* or the *pyramidal weighting function* as introduced in App. D.3, are also possible. In contrast to the Gaussian of eq. (2.67), they possess compact support, see Fig. 2.18 a).

The steps of calculation for a model evaluation are illustrated in Fig. 2.18 b) for a one dimensional input. All measured data points  $(u(i), y(i))$  for model training are stored as shown in the top most plot. Each output is weighted with regard to a normalised weighting function  $K_{\mathbf{h}}(u, u(i))$  (middle plot) and summed up to a model output as shown in the lower plot.

As the entire training dataset needs to be stored and regarded for model evaluation, the required memory and the computational effort depends on the dimensionality of the model input and the number of measurements. To reduce the number of kernel functions, a selection can be performed as is done for example for the identification with *support vector machines* [176, 179]. Due to its good adaption to measurement data, the Kernel model can be used as universal approximator, if sufficient measurements are available. A drawback is however its missing model structure and the dependency of computing time and memory demand on the number of measurements. This complicates an usage for real time applications as is discussed in more detail in Sect. 6.1.



**Figure 2.18:** a) Gaussian, Epanechnikov and pyramidal activation function for an usage in the Kernel model as  $K_h(\mathbf{u})$  and b) calculation of the model output for a Gaussian activation function for a 1-dimensional Kernel model.

## 2.6 Summary

The experimental model structures look-up tables, polynomial models, LOLIMOT, LOPOMOT and a Kernel model are presented. They can be regarded as black-box models if no a-priori information is applied or as grey box models if some structural information is included into the models.

Look-up tables are easy to interpret, which is why these are common used model structures in automotive applications. The model parameters can be identified by an explicit least-squares, a recursive least-squares or an approach with an intermediate model. Depending on the excitation, it might be necessary to introduce a regularisation into the identification. The presented regularisation penalises the curvature of the model and the degree of penalisation is adjusted by the regularisation parameter  $\lambda$ . The size of a suited regularisation parameter depends on the process and the number of measurements. Therefore, a formulation of a regularisation parameter is presented that is independent to the number of measurements. Such a formulation can also be applied to preserve the smoothness for a recursive adapted look-up table.

Since look-up tables are mostly limited to two input variables, combinations of two-dimensional look-up tables are presented to map higher dimensional relations. These are, the combination of a look-up table with a complex model structure, the additive combination of several look-up tables, the multiplicative combination of several look-up tables and nested combinations of look-up tables. Besides the functional approximation of a process, look-up tables can also be applied for control tasks. For some control tasks, the inversion of a look-up table is necessary. Therefore, a monotony characteristic is formulated for a look-up table, which needs to be ensured at model identification. The inverse can then either be determined by an inverse evaluation of the look-up table or by an approximation of the inverse with another look-up table.

Polynomial models are motivated by the Taylor series of a function. Continuous functions can be approximated locally arbitrary well by increasing the order of the Taylor series. With an increase of the order, the number of regressors increases exponentially, why a selection of regressors is presented. This selection determines the significant regressors and cancels redundant regressors. The applied criterion of fit is Mallows'  $C_p$ -statistic and the heuristic selection algorithm is a combination of a forward selection, backward selection and replacement of regressors. If a polynomial model structure is determined, the measurement design can be adjusted to this structure by the Design of Experiments. This is especially beneficial if a model structure is derived from the selection of regressors.

The multi-model approach LOLIMOT is presented, which enables to map different process properties within one model. LOLIMOT uses local linear models and composes them by means of Gaussian activation functions to a global model output.

A multi-model approach applying polynomial models is introduced by LOPOMOT. This structure combines the positive properties of LOLIMOT and the adaptive polynomials. The selection of regressors for the polynomial model is therefore further developed to a local formulation. Based on the local criterion of fit, a global criterion to rate the model quality is derived to select the best model partition. Because of the increased local model complexity and the local selection of regressors, LOPOMOT requires less model partitions and less parameters to achieve comparable model qualities to LOLIMOT. It is especially superior if the number of measurements is relatively small and if not all model parameters are excited by the training data.

The Kernel model is a data driven approach and can be regarded as universal approximator. The output of the Kernel model is determined by the weighted sum of the measured outputs from the training data. The weights are determined by a kernel function, which depends on the measured inputs from training data and a bandwidth parameter. The quality of the model mainly depends on the coverage of the input space with measurements. The only parameters that are identified are the bandwidth parameters of the weighting function.

The suitability of one of these model structures depends on the field of application. For the combustion model presented in the following, the LOPOMOT model structure is applied. A comparison of the model structures with regard to computing time, model quality, required memory and interpretability is given in Chap. 6.

---

## 3 Static and Dynamic Combustion Models

---

Mean value models are presented for the combustion outputs  $\text{NO}_x$ , soot and engine torque in this chapter. The dynamic influences on the combustion process are separated from the stationary influences. Various model inputs and output transformations are discussed and a global-local model structure is introduced. The model performances are evaluated on stationary and dynamic measurements for training and validation data. Stationary and dynamic emission characteristics are analysed using model simulations.

---

The calibration of modern combustion engines requires detailed models for the engine torque and the raw emissions. These models are in the following denoted as combustion models. Depending on the field of application, there are different model structures appropriate to describe the combustion process. Model structures can in general be divided into *crank angle synchronous models* and *mean value models*. The former are primarily white-box models based on physical laws and chemical reaction equations, while the latter are usually black-box models based on experiments.

### Crank angle synchronous models

Crank angle synchronous models can be differentiated by their level of detail [101]. The least complex models are *zero-dimensional* models. These models are based on simple thermodynamic laws or on empirical relations, such as the Vibe function [178] or the Mixing Controlled Combustion approach [30]. Zero-dimensional models are fast to calculate, but do mainly describe the rate of heat release and not the formation of emissions. To simulate the emissions, the more complex *phenomenological* models are required. Phenomenological models describe the emissions by dividing the combustion chamber into small packages [56] and applying chemical equations, such as the Zeldovich mechanism [54]. The computing time of these models can be up to several minutes for a single combustion, in contrast to the zero-dimensional models, which can be simulated in less than a second.

Applying *computational fluid dynamic (CFD)* models, an even more detailed understanding of the combustion process can be obtained. These are the most complex models and are mainly based on the Navier-Stokes equation, a spray model and a model for the turbulences. CFD models are simulated by a finite volume approach and the computation of a single combustion process can take up to several hours. An overview of CFD models is given in [101]. CFD models are well suited to design the combustion chamber, the injectors and the injection system. They are however too complex for a model based optimisation of the engine control. Furthermore, the accuracy of the emission simulation is, despite the high computational effort, often not satisfactorily. This is due to

the stochastic effects in emission formation and the high oxidation rate in the case of soot emissions. Up to 99 % of the produced soot is oxidated again to the end of the combustion process such that high relative errors can occur [159]. Therefore, experimental mean value models are applied in the following to model the combustion process.

### Mean value models

For a mean value model, the combustion process can be regarded as a cyclic batch process. Such a batch process depends only on the initial states and the boundary conditions of combustion. The initial states are the states of the air path at intake valve closing and the boundary conditions are the operation point of the engine and the characteristic values of the fuel injection system. The operation point is defined by the engine speed and the injection quantity. The characteristic values of the fuel injection system are among others the injection angle, the rail pressure and the splitting of injections. The various model inputs are regarded in detail in the subsequent section. In the following, the design of the mean value model is discussed.

Since the combustion engine operates a large part of its time in transient operation, the identified model needs to be able to simulate the engine dynamics. One approach is to model the combustion process by a discrete difference equation and introduce the dynamics as external dynamics, see [13, 48, 148]. Alternatively, a state space model can be applied, see [21, 182]. These models are identified by a dynamic excitation of the engine [189]. Such an excitation has the drawback that the dynamic data is naturally affected by measurement dynamics. Depending on the test bed, these dynamics can be significant especially for the emission measurements.

The location of emission measurement is typically downstream from the outlet valves, which leads to a dead time and to mixing effects in the exhaust pipe. Furthermore, the emission sensors themselves possess a dynamic. These dynamics are summarised to a measurement dynamic, which is undesired for a functional description of the process. A detailed investigation of the measurement dynamics for the  $\text{NO}_x$ , soot and opacity measurements on the test bed of the *Institute of Automatic Control and Mechatronics* is presented in [195]. Excitation signals are described and a second order time lag model with dead time is identified,

$$G(s) = \frac{1}{(T_1s + 1)(T_2s + 1)} e^{-T_d s}. \quad (3.1)$$

The time lags are due to mixing effects and sensor dynamics and the dead time is caused by the travelling time through the exhaust pipe and the flexible tubes to the measurement devices.  $\text{NO}_x$  is measured by an NGK  $\text{NO}_x$  sensor, opacity by an AVL 439 Opacimeter and soot by an AVL 483 Micro Soot sensor. For each of these measurements, model parameters for eq. (3.1) are identified in dependency on the volume flow rate of the exhaust  $\dot{v}_{\text{exh}}$ , see Tab. 3.1. Similar model structures to model the measurement dynamics are presented in [3, 8].

It is principally possible to invert and remove the measurement dynamics from the dynamic data. However, the dependencies of the dead time and the time lag on the volume flow rate complicates

**Table 3.1:** Measurement dynamics for the  $\text{NO}_x$ , soot and opacity measurements for the test bed at the Institute of Automatic Control and Mechatronics [195]. Dead times  $T_d$  and time constants  $T_1$  depend on the volume flow rate of the exhaust  $\dot{v}_{\text{exh}}$ , with  $\dot{v}_{\text{exh}} \in \left[0.01 \frac{\text{m}^3}{\text{s}}, 0.12 \frac{\text{m}^3}{\text{s}}\right]$ . The time constants  $T_1$  are limited to  $T_1 \geq 0$  s.

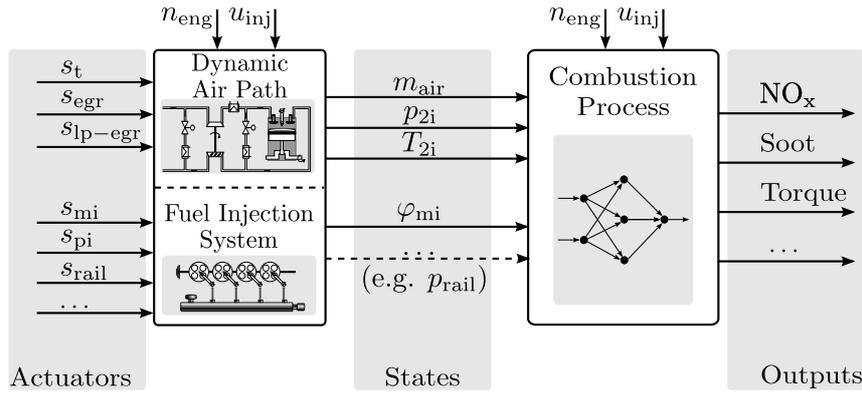
	$T_d$	$T_1$	$T_2$
$c_{\text{nox}}$	$0.21 \text{ s} + \frac{0.0019 \text{ m}^3}{\dot{v}_{\text{exh}}}$	$0.23 \text{ s} - 2.45 \frac{\text{s}^2}{\text{m}^3} \dot{v}_{\text{exh}}$	0.45 s
$c_{\text{soot}}$	$1.95 \text{ s} + \frac{0.0169 \text{ m}^3}{\dot{v}_{\text{exh}}}$	$1.01 \text{ s} - 4.56 \frac{\text{s}^2}{\text{m}^3} \dot{v}_{\text{exh}}$	0.0 s
$c_{\text{opa}}$	$0.20 \text{ s} + \frac{0.0116 \text{ m}^3}{\dot{v}_{\text{exh}}}$	$0.09 \text{ s} - 0.36 \frac{\text{s}^2}{\text{m}^3} \dot{v}_{\text{exh}}$	0.0 s

the inversion. Furthermore, an inversion of the time lag leads to an differentiating behaviour and therefore increases the noise.

In contrast to the emissions, the produced torque is directly transferred by the mechanical parts of the engine to the drivetrain, where it is measurement by the torque sensor at the asynchronous machine. Hence, there is no travelling time for the torque measurement and the dynamic effects of the drivetrain and the sensor itself can be neglected. There is however an observable dynamic deviation for the torque measurement for variations in engine speed. These deviations are caused by the rotation speed control of the asynchronous machine, since variations in engine speed are adjusted by a lower or higher load of the asynchronous machine to accelerate or decelerate the engine. This deteriorates the torque measurement on the asynchronous machine for the time of adjustment. Hence, the measured torque is too high respectively too low during negative respectively positive transients in engine speed.

To avoid these measurement dynamics, especially for the emission measurements, stationary model structures are applied for the combustion models. These models are then calibrated by stationary measurements, which are not affected by measurement dynamics. Although stationary model structures are applied, these are capable to cover the relevant dynamics, since the combustion process can be regarded as a batch process for a mean value model. The outcome of the batch process depends only on the initial states and the boundary conditions of combustion. Hence, if the model inputs are measured dynamically fast enough, the dynamics in emission formation are introduced by the dynamics of the air path, whereas the model structure of the emission model remains stationary. Figure 3.1 shows the block diagram for this model structure with the inputs of the combustion process being taken from the dynamic air path and the fuel injection system.

The main dynamics in this model structure are introduced by the turbocharger and the air path, including the exhaust gas recirculation (egr) system, with time constants up to 2.0 s. In contrast, the only dynamic contribution of the fuel injection system is given by the common rail. The injection characteristics, such as the crank angle of main injection  $\varphi_{\text{mi}}$ , can be adjusted cyclewise and possess therefore no significant dynamics. Since the rail pressure is relatively fast compared to the dynamics



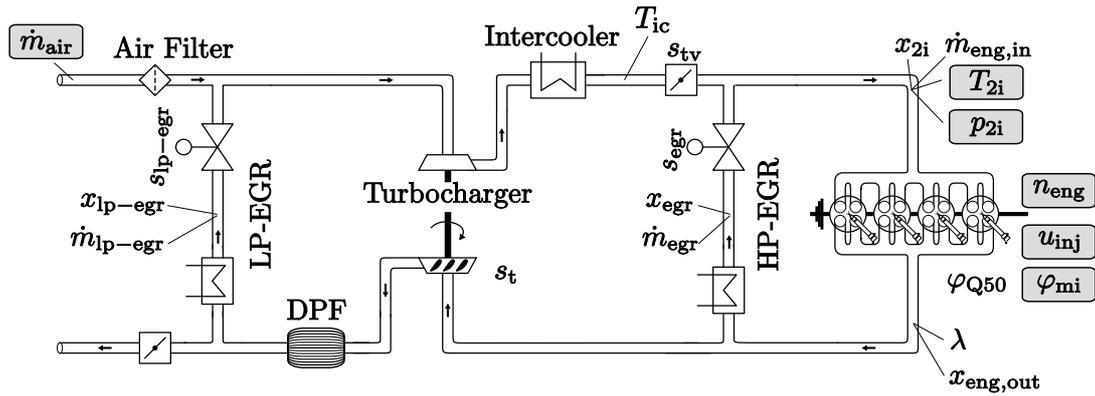
**Figure 3.1:** Block diagram of the dynamic air path, the fuel injection system and the stationary combustion process. Each of the blocks can be modelled and run separately if model inputs are provided. The models can also run in combination for an overall engine simulation. A model of the combustion process consists of several MISO models for the emissions and the engine torque. The main dynamics are introduced by the air path, why a dynamical model structure is required here. The dynamics of the fuel injection system and the combustion process are neglected for the applied mean value model.

of the air path, it is neglected for the presented model. It can however easily be integrated in the model structure if a more detailed model is required.

For an overall engine simulation all parts in Fig. 3.1 need to be modelled. The air path needs to be modelled by a dynamic model structure, which can either be a black-box model with external dynamics as presented in [59, 181] or a semi-physical model with lumped parameters as shown in [43, 131]. For the applications presented in this work, the semi-physical air path model is taken from [107] and is kindly provided by Matthias Mrosek. The model structure is shortly described in App. A. The structure of the air path with its main elements is depicted in Fig. 3.2. The highlighted variables are the applied inputs of the combustion model.

Regarding the injection system, it is assumed that the outputs follow their desired set points without any dynamics. Hence, no model is required here. The connection of the fuel injection system model, the air path model and the model of the combustion process enables a dynamic overall engine simulation from the engine actuators to the combustion outputs. If the states of the air path are available from measurements, as is the case during engine operation, the combustion model can also be utilised separately.

In the following, the combustion model is regarded in more detail. There are several input combinations feasible for the combustion model, also depending on the regarded output. Possible inputs and several output transformations are presented. Then, the model structure is introduced in Sect. 3.2. The measurement design for model identification is discussed in Sect. 3.3. Results for training and validation data for stationary and dynamic measurements are shown in Sect. 3.4. In Sect. 3.5, characteristics of the emission formation are analysed applying an intersection plot tool, which



**Figure 3.2:** Air path structure with a turbocharger, a low-pressure egr (LP-EGR) and high-pressure (HP-EGR) system. Fresh air enters the system at the upper left corner and leaves the system at the lower left. Common variable notations are applied and inputs of the combustion model are highlighted.

enables a visualisation of the high dimensional relations. Finally, the transient engine operation is investigated by a dynamic overall engine simulation in Sect. 3.6.

## 3.1 Model inputs and outputs

There is a variety of input combinations presented in the literature to model the combustion process. This variety mainly concerns the inputs coming from the air path system. Several input combinations can be transformed into each other and are redundant from an analytic point of view. However, due to the non-linear dependencies and also depending on the field of application, some input combinations are better suited than others. In the following, the applied inputs are introduced and common alternatives are discussed. Thereafter, several output transformations are presented.

### 3.1.1 Model inputs

The model inputs can be classified in inputs coming from the *operation point*, from the *air path states* and from the *injection characteristics*.

#### Operation point

The engine operation point is applied as global model input. It is defined by the *engine speed*  $n_{\text{eng}}$  and the *injection quantity*  $u_{\text{inj}}$ . These states are implicitly respectively explicitly given by the drivers request. The engine speed is implicitly given as it results from the engine speed and the selected gear. The injection quantity results directly from the position of the acceleration pedal, but can contain corrections from the ECU.

Alternative to the injection quantity, the engine torque is often applied to describe the engine operation point. The use of the engine torque is however not recommended here, as it regarded as an output of the combustion process which is also effected by the air path states and the injection characteristics.

### Air path states

From the air path states, the most common used model inputs are the *air mass flow rate*  $\dot{m}_{\text{air}}$  and the *pressure at intake*  $p_{2i}$ . The absolute value of the air mass flow rate in kg/h has the drawback that it depends on the engine speed  $n_{\text{eng}}$ . The higher the engine speed, the higher the flow rate through the engine. To achieve a better interpolation over the engine speed, it is transformed into the *air mass per cycle*  $m_{\text{air}}$  in mg/cyc [203],

$$m_{\text{air}} = \frac{\dot{m}_{\text{air}}}{2n_{\text{eng}}} \frac{10^6}{60}. \quad (3.2)$$

Since the *temperature at intake*  $T_{2i}$  affects the cylinder charge via the gas density, it is also included as model input. A drawback of the temperature is the relatively slow temperature sensor dynamics. Since the intake temperature is mainly affected by the proportion of hot recirculated exhaust gas and cold fresh air, the sensor dynamics can partially be avoided by calculating the intake temperature via a combination of the temperatures of the exhaust gas mass flow rate and the fresh air mass flow rate,

$$T_{2i} \approx \frac{(\dot{m}_{\text{air}} + \dot{m}_{\text{lp-egr}}) T_{\text{ic}} + \dot{m}_{\text{egr}} T_{\text{egr}}}{(\dot{m}_{\text{air}} + \dot{m}_{\text{lp-egr}} + \dot{m}_{\text{egr}})}. \quad (3.3)$$

$T_{\text{egr}}$  is the measured egr temperature,  $T_{\text{ic}}$  is the measured temperature after intercooler,  $\dot{m}_{\text{egr}}$  is the modelled egr mass flow rate and  $\dot{m}_{\text{air}}$  the fresh air mass flow rate. If a low-pressure egr system is utilised, the low-pressure egr mass flow rate  $\dot{m}_{\text{lp-egr}}$  is added to  $\dot{m}_{\text{air}}$ , otherwise  $\dot{m}_{\text{lp-egr}} = 0$ .

A common used model input is the *egr-rate*  $r_{\text{egr}}$  [13, 43, 52]. The egr-rate describes the proportion of fresh air and recirculated gas and is defined as

$$r_{\text{egr}} = \frac{\dot{m}_{\text{egr}} + \dot{m}_{\text{lp-egr}}}{\dot{m}_{\text{air}} + \dot{m}_{\text{egr}} + \dot{m}_{\text{lp-egr}}}. \quad (3.4)$$

A drawback of the egr-rate is that it does not regard the quality of the exhaust gas. The quality of the exhaust gas can be given by the proportion of oxygen in the exhaust gas. It can therefore be described by the *gas oxygen composition*  $x$ , which is generally defined as

$$x = \frac{x_{\text{air,fresh}}}{x_{\text{gas,total}}}. \quad (3.5)$$

The values of the gas compositions range from 1 for fresh air with 20.9% oxygen to 0 for burnt gas containing 0% oxygen. There are dynamics in the exhaust gas quality. These dynamics are especially apparent if a low-pressure egr system is utilised, since the travelling times are significant longer compared to a high-pressure egr system. These dynamics are covered by the gas composition

$x$ , but are not covered by the egr-rate  $r_{\text{egr}}$ , why the gas composition is recommended here. The gas composition dynamics are investigated in detail in [107] and are also shortly described in App. A.2.

Possible gas composition inputs are the *gas composition at intake*  $x_{2i}$  and the *gas composition after combustion*  $x_{\text{eng,out}}$ . For stationary conditions, the gas composition at intake is defined as

$$x_{2i} = \frac{m_{\text{air}} + x_{\text{egr}}m_{\text{egr}} + x_{\text{lp-egr}}m_{\text{lp-egr}}}{m_{\text{eng,in}}}, \quad (3.6)$$

with  $m_{\text{eng,in}}$  being the total gas mass per cycle and  $x_{\text{egr}}$  and  $x_{\text{lp-egr}}$  being the gas compositions of the high-pressure egr and low-pressure egr respectively. The egr masses  $m_{\text{egr}}$  and  $m_{\text{lp-egr}}$  are estimated by the flow equation for compressible fluids through an orifice, see e. g. [43]. The gas compositions  $x_{\text{egr}}$  and  $x_{\text{lp-egr}}$  are simulated by the air path model, see App. A. The total gas mass per cycle  $m_{\text{eng,in}}$  can either be modelled by using the volumetric efficiency of the engine or by using the sum of gas masses  $m_{\text{eng,in}} = m_{\text{air}} + m_{\text{egr}} + m_{\text{lp-egr}}$ . For a dynamic determination of  $x_{2i}$ , the dynamics of the egr paths and the intake need to be considered. The air path is therefore modelled by a lumped parameter approach with several storages, see App. A.

Given  $x_{2i}$ , the gas composition after combustion can be calculated by

$$x_{\text{eng,out}} = \frac{m_{\text{air,eng}} - m_{\text{air,burnt}}}{m_{\text{gas,total}}} \quad (3.7)$$

$$= \frac{x_{2i}m_{\text{eng,in}} - u_{\text{inj}}\rho_{\text{Diesel}}L_{\text{st}}}{m_{\text{eng,in}} + u_{\text{inj}}\rho_{\text{Diesel}}}, \quad (3.8)$$

with  $L_{\text{st}}$  being the stoichiometric ratio of Diesel and  $\rho_{\text{Diesel}}$  being the density of Diesel. Substituting eq. (3.6) into eq. (3.8) gives

$$x_{\text{eng,out}} = \frac{m_{\text{air}} + x_{\text{egr}}m_{\text{egr}} + x_{\text{lp-egr}}m_{\text{lp-egr}} - u_{\text{inj}}\rho_{\text{Diesel}}L_{\text{st}}}{m_{\text{eng,in}} + u_{\text{inj}}\rho_{\text{Diesel}}}. \quad (3.9)$$

For stationary conditions  $x_{\text{eng,out}} = x_{\text{egr}} = x_{\text{lp-egr}}$  and hence eq. (3.9) simplifies for stationary conditions to

$$x_{\text{eng,out}} = \frac{m_{\text{air}} - u_{\text{inj}}\rho_{\text{Diesel}}L_{\text{st}}}{m_{\text{air}} + u_{\text{inj}}\rho_{\text{Diesel}}}. \quad (3.10)$$

The gas composition after combustion can also directly be related to the *air to fuel ratio (after combustion)*  $\lambda$  by

$$x_{\text{eng,out}} = \frac{(\lambda - 1)L_{\text{st}}}{\lambda L_{\text{st}} + 1}, \quad (3.11)$$

see App. A.1. The air to fuel ratio is measured in the exhaust pipe and with eq. (3.11) this measurement can be applied to determine the gas compositions of the recirculated exhaust gas  $x_{\text{egr}}$  and  $x_{\text{lp-egr}}$ , see [18]. Thus, the uncertainties of the air path model are reduced by measurements. The measured air to fuel ratio can also directly be applied as model input, but since the measurement is located in the exhaust pipe after the combustion, the measured value is not well suited as model input. Hence, eq. (3.8) shall be used to determine  $x_{\text{eng,out}}$  for the emission model.

The gas composition at intake  $x_{2i}$  is applied in [17] as model input. In [7] it is applied together with the air to fuel ratio  $\lambda$  and in [5] together with the gas composition after combustion  $x_{\text{eng,out}}$ . Since soot mainly depends on the local availability of oxygen in the last phase of combustion, the gas composition after combustion  $x_{\text{eng,out}}$  is more informative for the soot emissions, than the air mass per cycle  $m_{\text{air}}$ . It is therefore applied in the following for the soot model, while the air mass per cycle is applied for the  $\text{NO}_x$  and engine torque model. The applied model inputs and outputs are again summarised in Tab. 3.2.

Alternative inputs to the above presented are applied by [47, 150]. They use the *actuator positions of the turbocharger*  $s_t$  and the *egr valve*  $s_{\text{egr}}$  to model the emissions. The dynamics of the air path are then included in the model, why a dynamic structure is required. Then dynamic measurements need to be applied which have the drawback that also the measurement dynamics are included in the models.

### Injection characteristics

The usage of a common rail systems enables the variation of the rail pressure and a splitting of the injection in pilot, main and post injections. Modern common rail systems can built up rail pressures up to 2000 bar and provide up to eight injections per cycle [90, 147]. Future systems will even reach rail pressures up to 3000 bar [156]. The injections can be adjusted cyclewise and possess therefore no dynamics regarding the mean value model. On the other hand, the setting of the rail pressure possesses a dynamic, which mainly depends on the operation point. Since this dynamic is relatively fast, with an average time constant of 0.022 s [110], it is neglected here. The rail pressure may however be included as model input for future work. The modelling procedure is then analogue, but the model complexity increases, which increases the calibration effort of the model.

The splitting and timing of injections influence the combustion process cyclewise and can therefore be adjusted from one batch process to another without any dynamics. Since the injection pattern is usually defined over the engine operation point, it is implicitly included in the engine operation point. Even if the injection pattern is adjusted with regard to the air path states, the influences are still covered by these inputs. It is therefore not necessary to include the splitting and timing of injections exclusively in a combustion model running on an ECU. On the other hand, if the combustion model shall be used for optimisation of the injection pattern, the influences of these quantities are of interest.

If the injection pattern shall be regarded, the *crank angles of the  $i$ -th pilot injections*  $\varphi_{pi,i}$  and the *quantities of the  $i$ -th pilot injections*  $q_{pi,i}$  need to be excited and included in the model structure. This leads to a significant increase in model complexity and therefore to an increasing calibration effort. Alternative to an optimisation via the inclusion into the combustion model, the injection pattern can also be optimised separately. This can either be done by an experimental analysis of the injection pattern [103, 113] or by applying models with detailed information about the combustion process itself, such as CFD models [57, 77, 101].

Due to the increasing complexity with additional model inputs, only the *crank angle of main injection*  $\varphi_{mi}$  is included in the model structure and the pilot injections are disregarded. The crank angle of main injection is applied for the optimisation in Chap. 4 and also adjusted for transient engine operations in [196]. The pilot injections on the other hand are disabled. Analogue to the rail pressure, these can easily be included for future work if they should be regarded for a model-based optimisation. Then the measurement design needs to be adjusted for such a high dimensional input space, but the modelling procedures are analogue.

Alternative to  $\varphi_{mi}$ , the *crank angle of 50% mass fraction burnt*  $\varphi_{Q50}$  can be applied. It is more informative than  $\varphi_{mi}$ , but requires a measurement of the in-cylinder pressure. If no in-cylinder measurement is available,  $\varphi_{Q50}$  can either be modelled or  $\varphi_{mi}$  need to be applied. The model accuracies with  $\varphi_{Q50}$  are slightly better, but the differences are relatively small. In the following  $\varphi_{Q50}$  is applied if it is available, otherwise it is modelled applying  $\varphi_{mi}$  as model input.

### 3.1.2 Model outputs

The regarded outputs are the *engine torque* and the emissions *nitrogen oxide* ( $\text{NO}_x$ ) and *soot*. Other combustion outputs, such as the *crank angle of 50% mass fraction burnt*  $\varphi_{Q50}$  or the exhaust gas temperature can also be modelled by an analogue approach. In the following, transformations for the measured emission concentrations to emission mass flow rates are presented, since emission mass flow rates are better suited for an engine optimisation than the emission concentrations. Furthermore, a logarithmic transformation is motivated for the soot measurement, which enables a more accurate model for small magnitudes.

Emissions are usually measured as concentrations, but the cumulated emission masses are relevant for the legislative regulations, see Fig. 1.1. Emissions measured in parts per million  $c_x$  can therefore be transformed to a emission mass flow rate by

$$\dot{m}_x = c_x \frac{m_{\text{mol},x}}{m_{\text{mol,exh}}} \dot{m}_{\text{exh,out}} \quad \text{in} \quad \text{mg/s} = \text{ppm} \frac{\text{g/mol}}{\text{g/mol}} \text{kg/s}. \quad (3.12)$$

At this  $m_{\text{mol},x}$  is the molar mass of the measured emissions and  $m_{\text{mol,exh}}$  is the molar mass of the exhaust gas, which can be simplified to the molar mass of air (28.96 g/mol).  $\dot{m}_{\text{exh,out}}$  is the tailpipe gas mass flow rate. For  $\text{NO}_x$  the molar mass  $m_{\text{mol},x}$  is given as  $\text{NO}_2$  equivalent (46.01 g/mol), see [33].

The soot concentration  $c_{\text{soot}}$  is measured in  $\text{mg/m}^3$  and needs therefore be transformed by

$$\dot{m}_{\text{soot}} = \frac{c_{\text{soot}}}{\rho_{\text{exh,meas}}} \dot{m}_{\text{exh,out}} \quad \text{in} \quad \text{mg/s} = \frac{\text{mg/m}^3}{\text{kg/m}^3} \text{kg/s}, \quad (3.13)$$

where  $\rho_{\text{exh,meas}}$  is the density of the gas in the measuring cell. Applying these transformations, the emissions can be modelled as emission mass flow rates, which are more meaningful for an optimisation than the concentrations.

Alternatively the emissions can also be related to the engine power, which is given by

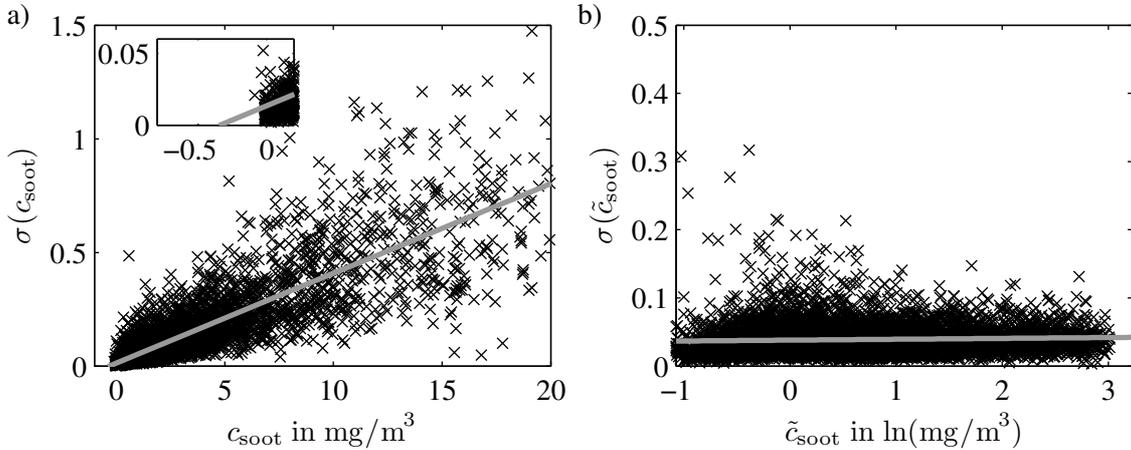
$$P_{\text{eng}} = M_{\text{eng}}\omega_{\text{eng}}, \quad (3.14)$$

where  $M_{\text{eng}}$  is the engine torque and  $\omega_{\text{eng}}$  is the engine speed in radian. The emission flow rates can then be related to the engine power expressed in kW such that the unit mg/kWh results.

Soot increases highly with a decrease of the air to fuel ratio  $\lambda$ . Simultaneously, the noise in soot measurement increases, which is due to the non-linear formation of soot and might also be due to a non-linear noise characteristic of the soot sensor. The variance of the soot measurement is estimated from the continuous data and plotted against the measured stationary magnitudes in Fig. 3.3 a). The variance increases linearly with the magnitude, which indicates a multiplicative error. Since a multiplicative error implies a variance of zero at  $c_{\text{soot}} = 0$ , a shift of  $c_{\text{soot}}$  is necessary. The shift of -0.34 is taken from the intersection point of the linear variance estimation (grey line) with the abscissa, where the estimated variance equals zero, see zoomed plot in Fig. 3.3 a). To be able to apply the least squares algorithm for parameter estimation, a logarithmic transformation is applied

$$\tilde{c}_{\text{soot}} = \ln(c_{\text{soot}} + 0.34). \quad (3.15)$$

This transforms the multiplicative error to an additive error as can be seen by the variance of the transformed soot  $\tilde{c}_{\text{soot}}$ , depicted in Fig. 3.3 b). The linear estimate of the standard deviation (grey line) can be regarded as constant over the magnitude, which indicates an additive error.



**Figure 3.3:** Standard deviations of the soot measurements over the magnitudes of the soot measurements. a) Standard deviations for the measured values  $c_{\text{soot}}$  and b) standard deviations for the transformed values  $\tilde{c}_{\text{soot}}$ .

The applied model inputs and outputs for the presented combustion model are summarised in Tab. 3.2. The first six inputs are used and two alternative applied inputs are stated below. The first alternative is the gas composition after combustion  $x_{\text{eng,out}}$ , which is more informative for the soot model and therefore applied instead of the air mass per cycle. The second alternative is the

**Table 3.2:** Modelled outputs and applied model inputs with units and measured intervals. The gas composition after combustion  $x_{\text{eng,out}}$  is applied for the soot model instead of the air mass per cycle  $m_{\text{air}}$ . The crank angle of 50% mass fraction burnt  $\varphi_{Q50}$  can be applied instead of the crank angle of main injection  $\varphi_{\text{mi}}$  if in-cylinder pressure measurements are available.

input	Inputs		Output	Outputs	
	interval	unit		interval	unit
$n_{\text{eng}}$	800 – 3000	rpm	$\dot{m}_{\text{nox}}$	0.1 – 159.0	mg/s
$u_{\text{inj}}$	0 – 30	mm <sup>3</sup> /cyc	$\dot{m}_{\text{soot}}$	0.0 – 5.1	mg/s
$m_{\text{air}}$	187 – 806	mg/cyc	$M_{\text{eng}}$	-53.9 – 133.5	Nm
$p_{2i}$	0.97 – 2.00	bar	$c_{\text{nox}}$	9.2 – 1345.4	ppm
$\varphi_{\text{mi}}$	-17.2 – 4.6	°CA after TDC	$c_{\text{soot}}$	0.0 – 48.7	mg/mm <sup>3</sup>
$T_{2i}$	15.0 – 137.7	°C	$\varphi_{Q50}$	8.9 – 21.1	°CA after TDC
(or) $x_{\text{eng,out}}$	0.06 – 0.80	-			
(or) $\varphi_{Q50}$	8.9 – 21.1	°CA after TDC			

crank angle of 50 % mass fraction burnt  $\varphi_{Q50}$ , which shows slightly better results than the crank angle of main injection  $\varphi_{\text{mi}}$ . It is applied if in-cylinder pressure measurements are available.

The suitability of a model input depends on the regarded output, the field of application and the availability of sensors. In general, the utilised inputs shall not be redundant. If an optimisation is performed with regard to the selected inputs, the model inputs need to be manipulable by the actuators and the number of inputs shall not exceed the number of actuators. Otherwise, one or more inputs need to be regarded as constants for the optimisation and the obtained results are then optimal with regard to the assumed constant values.

Since it is necessary to apply different model inputs for different model applications, a database of models can be stored from which a suited model is selected, see App. B.1. This database states the applied model inputs, the modelled output, the model quality and the number of model parameters. Models with alternative inputs can then easily be picked from this database.

## 3.2 Local and global model structures

Experimental combustion models can be distinguished by local and global model structures. *Local* models describe the combustion output with regard to a fixed engine operation point,

$$f_{\text{local}} = f_{n_{\text{eng}}, u_{\text{inj}}}(m_{\text{air}}, p_{2i}, \dots). \quad (3.16)$$

To identify such a local model, the local model inputs ( $m_{\text{air}}, p_{2i}, \dots$ ) are excited and the engine operation point is kept constant.

A *global* model allows the variation of the engine operation point and includes it in the identified model,

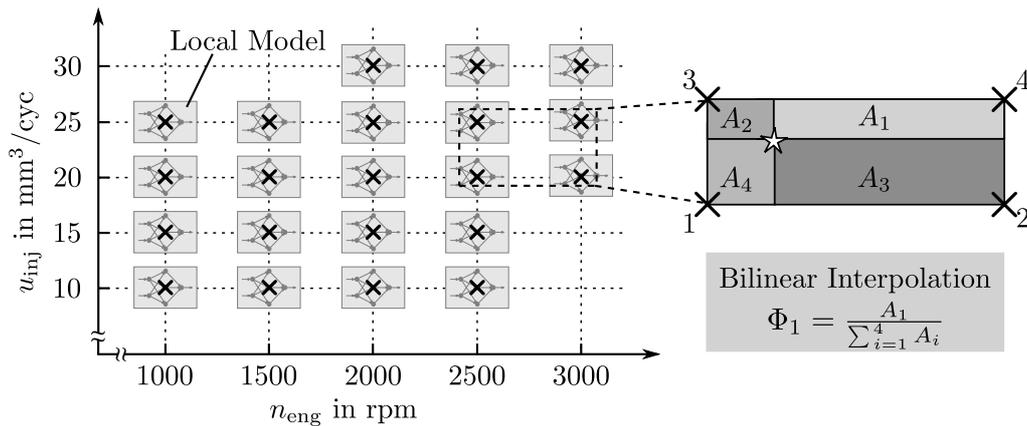
$$f_{\text{global}} = f(n_{\text{eng}}, u_{\text{inj}}, m_{\text{air}}, p_{2i}, \dots). \quad (3.17)$$

To identify a global model structure, all model inputs need to be excited simultaneously.

The model structure applied in the following describes the global model output by a composition of several local models. It is referred to as *global-local* model structure,

$$f_{\text{global-local}} = \sum_{j=1}^M \Phi_j(n_{\text{eng}}, u_{\text{inj}}) f_{n_{\text{eng}}, u_{\text{inj}}}(m_{\text{air}}, p_{2i}, \dots). \quad (3.18)$$

The local models are distributed over the engine operation points such that the relevant region of common test cycles is covered, see Fig. 3.4. The local model outputs are then composed by means of a bilinear interpolation to a global model output. Hence, if the model is evaluated in a model centre, only one local model is active. If the model is evaluated in between the model centres, the four surrounding models form the output by a weighted sum. The global weighting can therefore be interpreted as a look-up table structure, see Sect. 2.1.



**Figure 3.4:** Global-local model structure for the combustion models. The distribution of the local models over the engine operation points covers the relevant region of common test cycles. Local models are defined for fixed engine operation points. The global model output is determined by a bilinear interpolation of the local models. The bilinear interpolation can be interpreted as an area interpolation.

The local models are described as polynomials of order three. For these local polynomials, the selection algorithm from Sect. 2.2.1 is applied. This enables an adaption of the local models to the local non-linearity of the process. Since the pyramidal weighting function of the LOPOMOT model structure is equivalent to a bilinear interpolation for an equidistant lattice, the applied model structure can also be interpreted as LOPOMOT model with fixed model partitions. The resulting

model structure can then be written as

$$f_{\text{LOPOMOT}}(\mathbf{z}, \mathbf{u}) = \sum_{j=1}^M \Phi_j(\mathbf{z}) (w_1 x_1 + w_2 x_2 + \dots + w_n x_n) \quad (3.19a)$$

$$\text{with } \mathbf{z}^T = [n_{\text{eng}}, u_{\text{inj}}] \quad (3.19b)$$

$$\text{and } x_i \in \{1, u_1, u_2, \dots, u_4, u_1^2, u_1 u_2, \dots, u_4^3\} \quad (3.19c)$$

$$\text{and } \mathbf{u}^T = [m_{\text{air}} \text{ OR } x_{\text{eng,out}}, p_{2i}, \varphi_{\text{mi}}, T_{2i}]. \quad (3.19d)$$

The global model inputs are introduced as  $z$ -regressors to the model, being the engine operation point. The local regressors are selected from the set of potential regressors, eq. (3.19c). This set is derived from the local model inputs, stated in eq. (3.19d). For the soot model the gas composition after combustion  $x_{\text{eng,out}}$  is applied instead of the air mass per cycle  $m_{\text{air}}$ . The weighted sum of the local models, eq. (3.19a), gives the global model output.

It is also possible to apply alternative model structures than the utilised adaptive polynomials for the local models in eq. (3.18). A comparison of the model structures presented in Chap. 2 is given in Sect. 6.1 regarding the  $\text{NO}_x$  emission model. In all cases the global-local model structure is applied, which has several positive properties:

- The global-local model structure is suited for a local as well as for a global optimisation. For the local optimisation each local model is optimised separately and for the global optimisation the weighted sum of the local models is applied. The local optimisation is performed prior to the global optimisation, see Chap. 4. Therefore, the same model can be utilised for both optimisation steps.
- Drivability spaces can be defined for fixed engine operation points. A drivability space is applied to evaluate if an input combination is reachable in practice, which is required for the model-based optimisation. The local formulated drivability space can then be utilised for the local optimisation. The determination of drivability spaces is presented in the following section.
- The global-local model structure is fast to calculate and can easily be implemented on common ECUs. At most four local models need to be evaluated, which are then weighted by means of a bilinear interpolation. This interpolation is the same as for the widely used look-up tables.
- The local polynomial models are well suited for an optimisation, since the number of optima is defined by the polynomial order and, depending on the polynomial order and the optimisation problem, may also be determined analytically. Furthermore, a derivative can be determined analytically.
- The local models can easily be extended by further model inputs. These inputs can be included in the model structure either with or without interactions to the other inputs.
- The multi-model approach allows to model different properties for different operation points.

### 3.3 Measurement design

In the following, the measurement design for the identification of the global-local model structure is presented. Since a stationary model structure is applied, stationary measurements are used for model training and validation. Thus, the significant measurement dynamics of emission measurements are avoided. For the collection of stationary measurements, closed loop controls for the engine speed, the air path states and the crank angle of 50% mass fraction burnt are utilised. Each measurement point is hold for 15 s such that the dynamic effects of the system and the measurement dynamics are decayed. The measurements are averaged over the last 1.5 s such that noise effects are reduced.

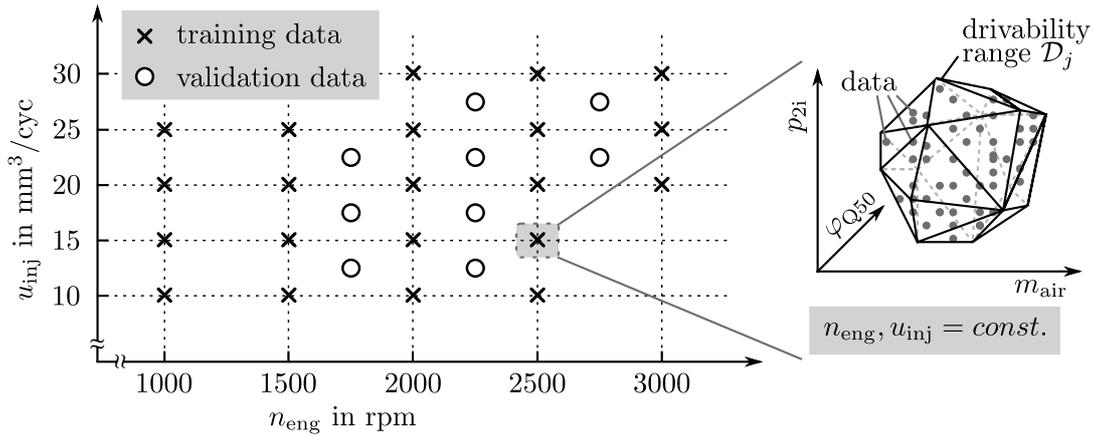
The choice of the holding time is based on the time constants of the system and the measurement dynamics. It is an upper bound with a safety distance such that all dynamic effects are decayed. After data collection, the measurement data is checked for outliers and if the holding time was chosen sufficiently long. This is a laborious task, which is mainly based on visual decision. It is also possible to analyse the data already at data collection automatically, which reduces the measurement time and increases the data quality. An example is the automatic detection of stationary states, instead of applying the upper bound holding time of 15 s. Such an automatic data collection is one of the objectives of a current research project at the Institute of Automatic Control and Mechatronics [191].

#### Distribution of measurements

The distribution of the measurement data over the engine operation points is depicted in the left part of Fig. 3.5. The training data is placed in the local model centres, compare to Fig. 3.4, and the validation data in between. For each operation point, several data points are distributed over the local drivability space, as is illustrated in the right part of Fig. 3.5. The drivability space of the  $j$ -th local model is denoted by  $\mathcal{D}_j$  and envelopes the reachable points. The drivability space may be determined with separate experiments or, as applied here, incorporated in the measurement design. A rough description of  $\mathcal{D}_j$  is derived from the semi-physical air path model, to determine the measurement design in each engine operation point. Then the measurements are performed and the final drivability space  $\mathcal{D}_j$  is defined by the convex hull surrounding the collected data points. Other structures than the convex hull can also be applied for  $\mathcal{D}_j$ , if the drivability space is not describable by a convex hull.

Measurements are performed on several days in the summer and several days in the winter with varying environmental conditions. The identified models can therefore be regarded as non sensitive to changes in environmental conditions. The measurement data consists of 4584 data points for model training and 2262 data points for model validation.

The measured combustion outputs are  $\text{NO}_x$ , soot, engine torque, crank angle of 50% mass fraction burnt, opacity, exhaust gas temperature, exhaust gas pressure and further emissions, such as CO or HC. For the presented optimisations the emissions  $\text{NO}_x$  and soot and the engine torque are of

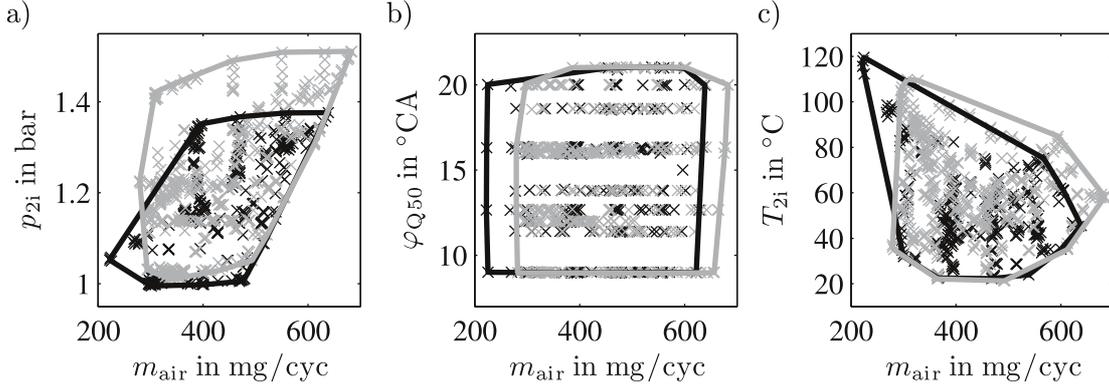


**Figure 3.5:** Distribution of the training and validation data over the engine operation points. The training data is collected in the model centres and the validation data in between. In each operation point the local drivability space  $\mathcal{D}_j$  is covered with measurements. The varied local inputs are  $m_{\text{air}}$ ,  $p_{2i}$ ,  $\varphi_{\text{mi}}$  and  $T_{2i}$ . There are 4584 stationary data points collected for model training and 2262 data points for model validation.

interest.  $\text{NO}_x$  is measured by an NGK  $\text{NO}_x$ -sensor and measurements for soot are taken by an AVL Micro Soot sensor. Both are measured as concentrations and are transformed to an emission mass flow rate by eq. (3.12) and eq. (3.13) respectively. The torque is measured by a torque shaft sensor and the crank angle of 50% mass fraction by in-cylinder pressure sensors in combination with an indicating system.

For an excitation of the local model inputs, measurements are equally distributed over the local drivability spaces, see Fig. 3.6. The drivability spaces, shown by surrounding lines, are well covered. For the presented distribution of measurements, no a-priori information about the model structure is applied. Since polynomial models are identified, a D-optimal design, as described in Sect. 2.2.2, would be advantageous. However, a D-optimal design would favour a polynomial model structure, compared to other model structures. This would bias the comparison of model structures presented in Chap. 6, why a rastered input distribution is applied. But, if the task is to calibrate the combustion model, a D-optimal design should be applied such that the measurement design is adapted to the model structure.

The manipulated local input variables are the air mass per cycle  $m_{\text{air}}$ , the intake pressure  $p_{2i}$ , the crank angle of 50% mass fraction burnt  $\varphi_{Q50}$  and the intake temperature  $T_{2i}$ . These inputs are adjusted by closed loop control with actuators for the guide vanes of the turbocharger  $s_t$ , for the high-pressure egr valve  $s_{\text{egr}}$ , for the low-pressure egr valve  $s_{\text{lp-egr}}$  and the crank angle of main injection  $\varphi_{\text{mi}}$ . Since no thermal conditioning system is available, the temperature  $T_{2i}$  is excited by the proportion of hot high-pressure egr and cold low-pressure egr. If exhaust gas is recirculated to the engine, this exhaust gas can either be taken from the low-pressure or the high-pressure egr system. Hence, the intake temperature  $T_{2i}$  is colder if the egr is taken from the cold low-pressure egr than if the same amount of hot high-pressure egr would be recirculated. Since the proportion of



**Figure 3.6:** Two-dimensional projections of training data on the input dimensions for the operation points: black:  $n_{\text{eng}} = 2000$  rpm,  $u_{\text{inj}} = 15$  mm<sup>3</sup>/cyc and grey:  $n_{\text{eng}} = 2000$  rpm,  $u_{\text{inj}} = 20$  mm<sup>3</sup>/cyc. The drivability spaces of the local models are shown by the convex hulls surrounding the measurements.

hot high-pressure to cold low-pressure egr exits only the temperature and not the gas composition, it is applied to manipulate the intake temperature.

### Avoidance of extrapolation

With the collected measurement data, a drivability space  $\mathcal{D}_j$  is defined in each engine operation point by the convex hull surrounding the local measurement data. The drivability space  $\mathcal{D}_j$  is then applied to check if the local model operates in extrapolation region. Therefore, a local measure is defined for the  $j$ -th local model,

$$\text{EM}_j(\mathbf{u}) = \begin{cases} 0 & : \mathbf{u} \in \mathcal{D}_j \\ 1 & : \mathbf{u} \notin \mathcal{D}_j. \end{cases} \quad (3.20)$$

These local measures are summed up to a global measure by means of the weighting function  $\Phi_j$ ,

$$\text{EM}(\mathbf{z}, \mathbf{u}) = \sum_{j=1}^M \Phi_j(\mathbf{z}) \text{EM}_j(\mathbf{u}). \quad (3.21)$$

The global extrapolation measure EM takes values between 0 and 1, depending on how many local models operate in extrapolation region and on their weighting functions  $\Phi_j$ . A threshold for extrapolation can then be chosen between 0 and 1. The threshold depends on the extrapolation performances of the local models, which on the other hand depend on the applied model structure and the distances between the local models.

The applied polynomial models are well suited for interpolation, but polynomial models tend to infinity in extrapolation region. To ease these asymptotic effects, the inputs and outputs of the local models are limited by a bounding box to their minimum or maximum observed training values.

To enable some extrapolation, the bounding box of the inputs can be extended by  $\pm 10\%$ . Thus, extrapolation is enabled without the risk of unreasonable values.

Since the drivability spaces differ for different engine operation points, see Fig. 3.6, the risk of extrapolation depends on the distances of the local models. A dense distribution of local models lowers the risk of extrapolation, but requires a high calibration effort. Hence, the required number of local models depends on the variations of the drivability spaces over the engine operation points. It further depends on the variation of the physical behaviour from one local model to another. The number of local models is therefore a trade-off between the calibration effort on the one hand and the extrapolation risk and model accuracy on the other hand.

There are 21 local models identified over the engine operation points with the equidistant resolution of  $\Delta n_{\text{eng}} = 500$  rpm and  $\Delta u_{\text{inj}} = 5$  mm<sup>3</sup>/cyc, see Fig. 3.4. Applying this distribution, the variation of the drivability space from one operation point to another is acceptable, as can be seen in Fig. 3.6. The coverage in the  $m_{\text{air}}$  dimension is improved by applying the air mass per cycle rather than the measured air mass flow rate, see [203]. Model qualities for training and validation data are presented in the following section.

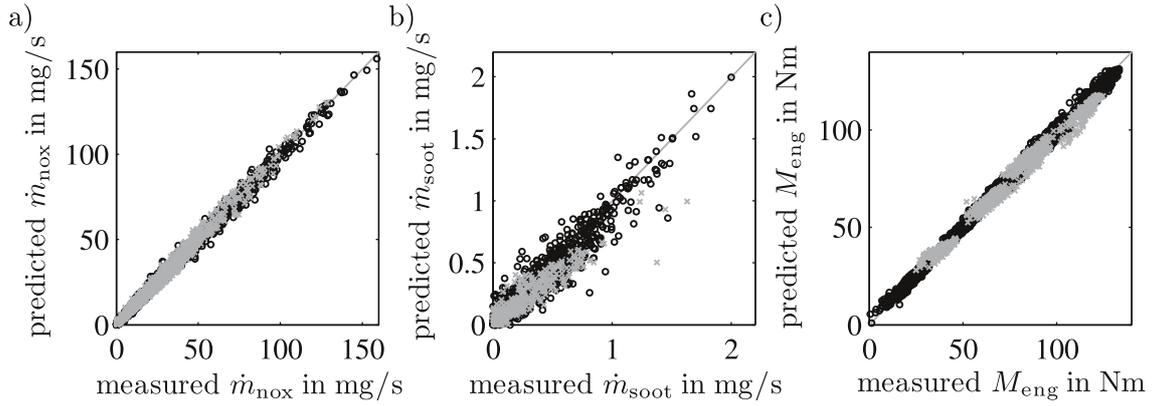
## 3.4 Training and validation results

### 3.4.1 Results from stationary data

The results for the model training and validation data, as described in Sect. 3.3, are shown in the measured versus predicted plots in Fig. 3.7. It can be seen that the NO<sub>x</sub> emissions are well modelled over all magnitudes. The variation around the bisecting line is low for training and validation data. For the soot emissions the deviation from the bisecting line increases with an increase of the magnitude. This is expected, since the low values are modelled more accurate for the logarithmic transformation than the high values. The torque model shows a clustering of the measurement points. This is due to the strong coupling of the engine torque to the injected fuel mass. Since the measurements are rastered in the injected fuel mass, the clustering appears in the output.

The qualities of the applied models are summarised in Tab. 3.3. Qualities are given in *coefficient of determination*  $R^2$  and as *root mean square error*  $RMSE$ . For the presented models, the crank angle of 50% mass fraction burnt  $\varphi_{Q50}$  is applied. If no in-cylinder pressure sensor is available,  $\varphi_{Q50}$  can be modelled or the crank angle of main injection  $\varphi_{\text{mi}}$  can be applied. The qualities of these models differ only slightly from the models applying  $\varphi_{Q50}$ , see App. B.2. For each of the 21 local models, 26 potential regressors are available, which makes 546 regressors for a global model. The numbers of selected regressors are stated in Tab. 3.3. Values are given for the modelled emission mass flow rates  $\dot{m}_{\text{nox}}$  and  $\dot{m}_{\text{soot}}$ , the modelled emission concentrations  $c_{\text{nox}}$  and  $c_{\text{soot}}$  and the modelled engine torque  $M_{\text{eng}}$ .

An extraction of the continuous regarded training data is shown in Fig. 3.8 for the engine operation point  $n_{\text{eng}} = 2000$  rpm and  $u_{\text{inj}} = 25$  mm<sup>3</sup>/cyc. It can be seen that each value is hold for 15 s



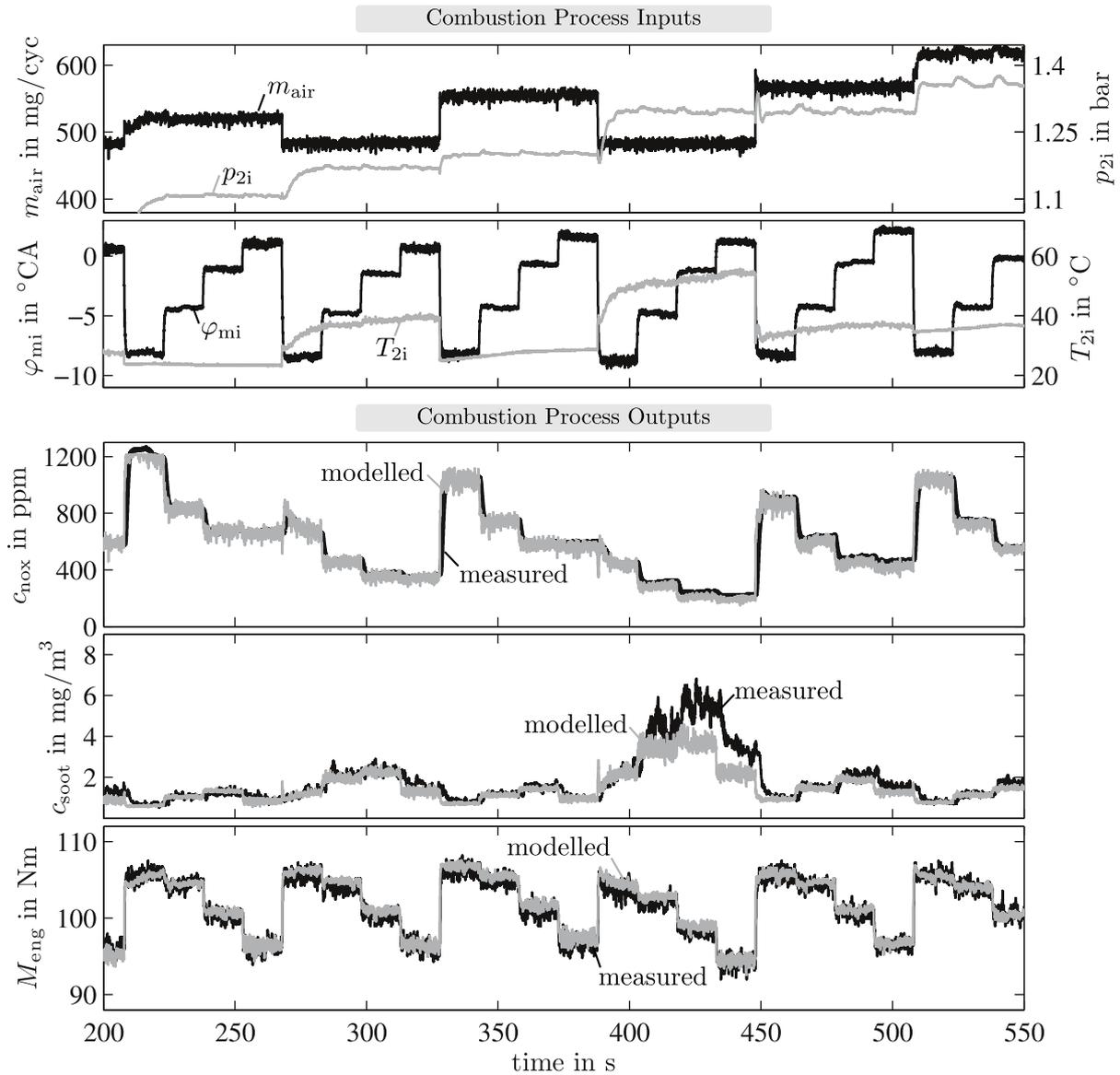
**Figure 3.7:** Measured versus predicted plots for the 4584 stationary training data points (black circles) and the 2262 stationary validation data points (grey crosses) for the emission mass flow rates of a)  $\text{NO}_x$  and b) soot and c) the engine torque.

**Table 3.3:** Model qualities for training and validation data for  $\text{NO}_x$ , soot and the engine torque. The emissions are either modelled as mass flow rates  $\dot{m}_{\text{NO}_x}$  and  $\dot{m}_{\text{soot}}$  or as concentrations  $c_{\text{NO}_x}$  and  $c_{\text{soot}}$ . The number of parameters is the sum of the number of selected regressors for the 21 local models.

Model	Training		Validation		# parameters
	$R^2$	RMSE	$R^2$	RMSE	
$\dot{m}_{\text{NO}_x}$	0.996	1.45 mg/s	0.988	2.09 mg/s	424
$\dot{m}_{\text{soot}}$	0.959	0.05 mg/s	0.903	0.06 mg/s	440
$M_{\text{eng}}$	0.997	1.52 Nm	0.981	2.83 Nm	233
$c_{\text{NO}_x}$	0.992	23.0 ppm	0.979	34.7 ppm	456
$c_{\text{soot}}$	0.939	0.87 mg/m <sup>3</sup>	0.907	0.97 mg/m <sup>3</sup>	439

until the main dynamic effects are decayed. From this continuous data, only one stationary point is extracted every 15 s, which is then applied for the model training and validation as shown in Fig. 3.7. Besides the measured inputs and outputs, the model outputs are plotted into the lower three plots. It can be seen, that a relative good model accuracy is obtained for the stationary behaviour. The dynamic variations are due to sensor dynamics and will be discussed in more detail in the following section.

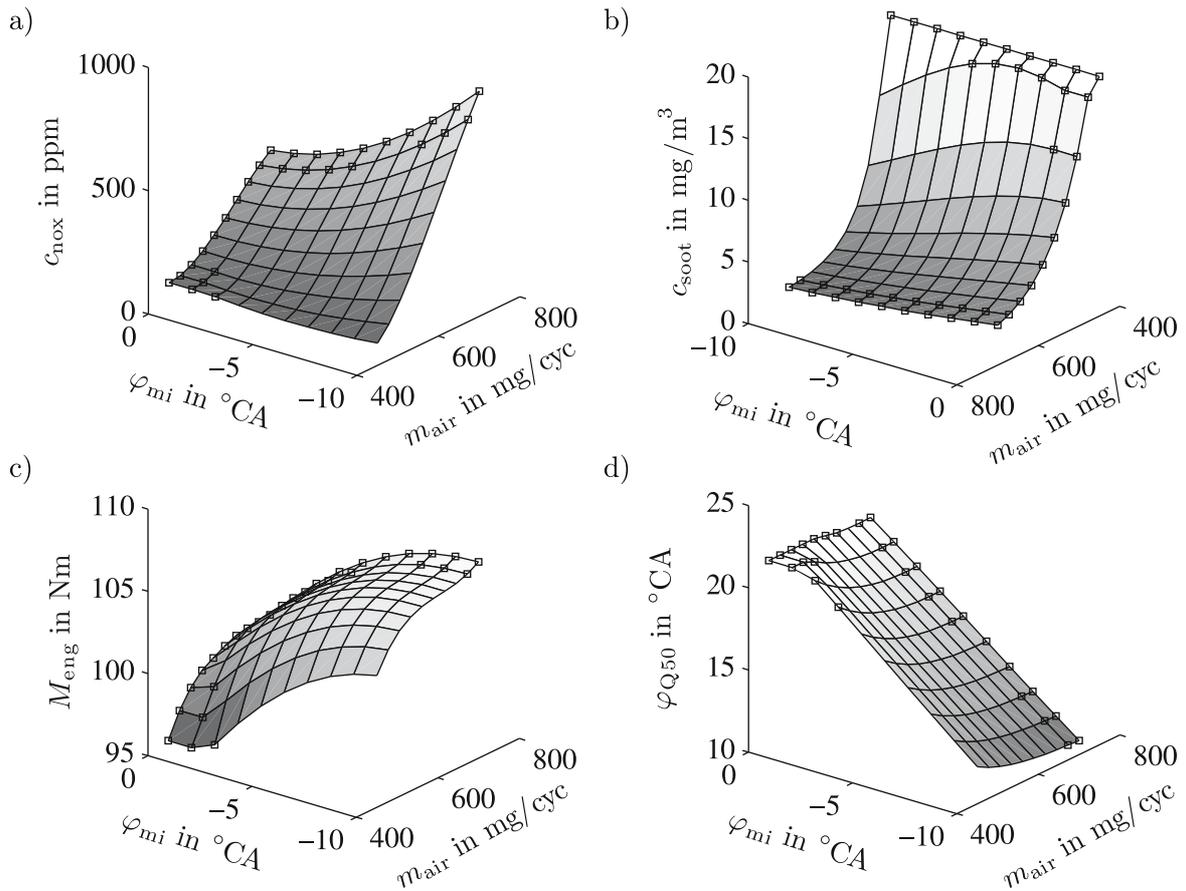
The stationary behaviour of the identified models is shown in Fig. 3.9 over the inputs  $m_{\text{air}}$  and  $\varphi_{\text{mi}}$  at the engine operation point  $n_{\text{eng}} = 2500$  rpm and  $u_{\text{inj}} = 25$  mm<sup>3</sup>/cyc. The model inputs  $T_{2i}$  and  $p_{2i}$  are kept constant. The black squares in the plot indicate the extrapolation regions of the models. Note that the axis are swapped for the plot of the soot model for a better visualisation. It can be seen that soot performs in opposite direction to the  $\text{NO}_x$  emissions and both emissions are more affected by the air mass per cycle than by the crank angle of main injection. On the other hand, a strong dependency on  $\varphi_{\text{mi}}$  is observable for the models for  $M_{\text{eng}}$  and  $\varphi_{\text{Q50}}$ .



**Figure 3.8:** Continuous measurements from which stationary points are extracted for model training and validation. Each point is hold for 15 s and the stationary point is averaged over the last 1.5 s. The presented measurements are from the engine operation point  $n_{\text{eng}} = 2000$  rpm and  $u_{\text{inj}} = 25 \text{ mm}^3/\text{cyc}$ .

### 3.4.2 Dynamic model results

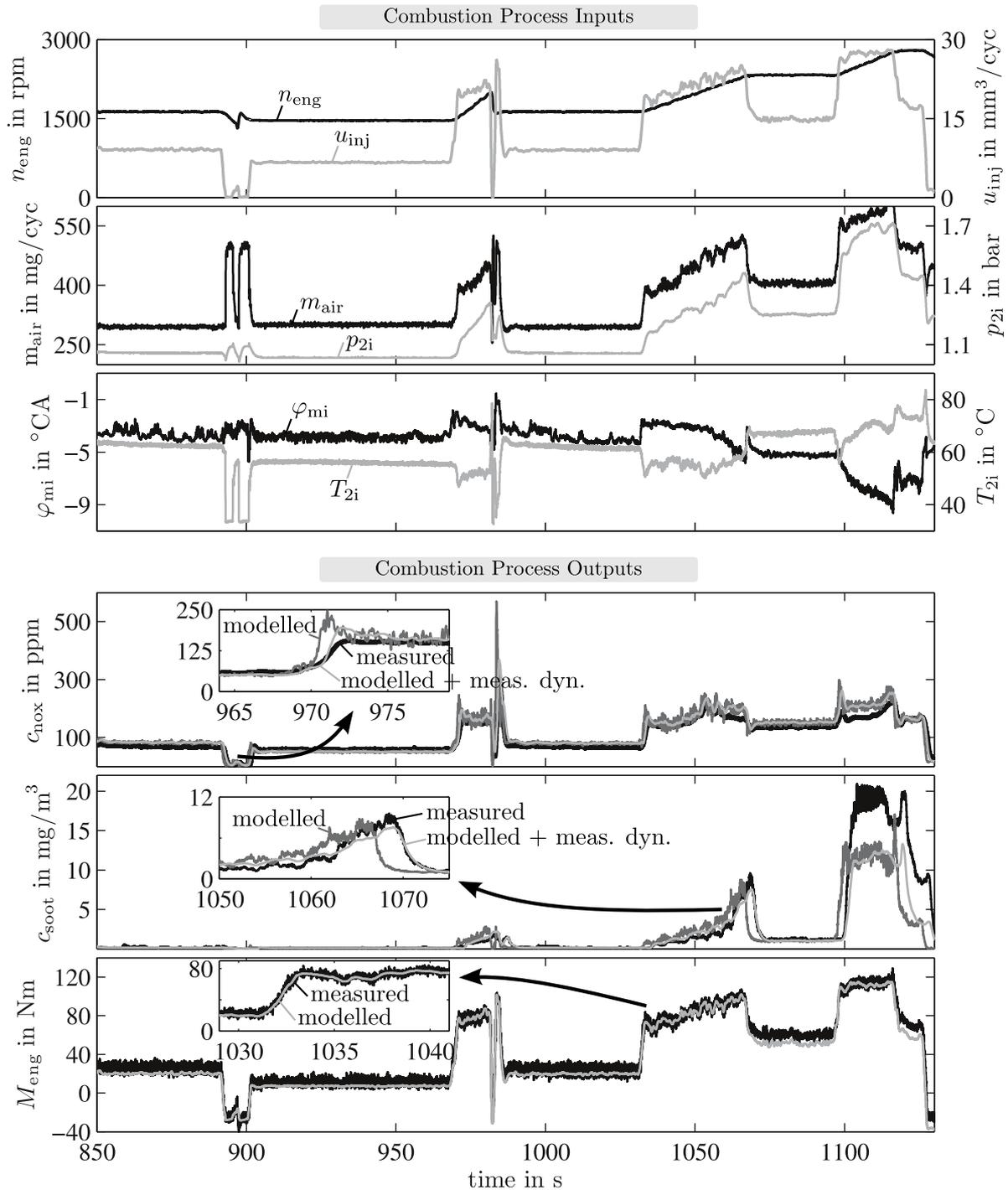
For a validation of the model dynamics, measurements of the extra-urban part of the *New European Driving Cycle (NEDC)* are presented in Fig. 3.10. The upper three plots show the measured model inputs and the lower three plots the measured and modelled outputs. The applied trajectories for engine speed and injection quantity (topmost plot) correspond to a test run on a roller test bench. The air path states are adjusted by closed loop controls (second plot) with the actuators of the turbocharger and the high-pressure egr valve. The desired values for the controllers are taken from a series calibration, as is also the value for the injection angle  $\varphi_{\text{mi}}$ . The low-pressure egr is disabled



**Figure 3.9:** Models for a)  $\text{NO}_x$  b) soot, c) torque and d) the crank angle of 50 % mass fraction burnt at the operation point  $n_{\text{eng}} = 2500$  rpm and  $u_{\text{inj}} = 25$  mm<sup>3</sup>/cyc. The other model inputs are kept constant at  $T_{2i} = 60$  °C and  $p_{2i} = 1.5$  bar. Note that the axis for soot are swapped for a better presentation of the look-up table. The black squares in the plot indicate the extrapolation region of the models.

for this measurement, why the intake temperature is not manipulated here. To compare the modelled emissions with the measured emissions, models for the emission concentrations  $c_{\text{NO}_x}$  and  $c_{\text{soot}}$  are applied.

The modelled emissions and the engine torque are presented with and without a model for the measurement dynamics in the lower three plots in Fig. 3.10. The stationary behaviour is well described for all outputs. The dynamical performance is investigated in detail in the zoomed plots. The torque model fits well to the measured dynamics. Since no sensor model is assumed for the torque model, this proves that the assumption of a batch process is well suited for a mean value model of the combustion process. It further proves that the stationary model structure is capable to model the system dynamics, which are introduced to the model by the dynamically fast measured inputs. For the torque model, the visible dynamics mainly result from dynamics in the injection quantity and are therefore included in this input.



**Figure 3.10:** Dynamical validation of the combustion models for the extra-urban part of the NEDC. Top three plots show the measured model inputs. Lower three plots show the measured emissions NO<sub>x</sub> and soot and the engine torque, together with their modelled values. Modelled emissions are also filtered with the models for emission measurement dynamics from eq. (3.1) with parameters from Tab. 3.1. For the torque model, no measurement dynamics are assumed. Measured outputs are black, modelled outputs are dark grey and modelled outputs with modelled measurement dynamics are light grey.

In contrast to the torque model, there are dynamic deviations between the modelled (dark grey line) and measured (black line) emissions, see the zoomed plots. The modelled emissions are well before the measured emissions and react considerably faster to changes in the inputs. These differences are mainly due to measurement dynamics. The influences of the travelling time can be seen for both, the  $\text{NO}_x$  and soot emissions. The shaping by the time constants is more pronounced for the  $\text{NO}_x$  measurement.

To compare the modelled emissions with the dynamic measured emissions, the models for the measurement dynamics, as in eq. (3.1), are applied to the modelled emissions (light grey). The parameters are taken from Tab. 3.1 with the volume flow rate of the exhaust  $\dot{v}_{\text{exh}}$  being calculated at each time step. The modelled emissions in connection with the modelled measurement dynamics fit well to the dynamic measured emissions. This proves that the observed dynamics in emission formation are mainly due to measurement dynamics and validates the emission models for dynamic applications. The application of the model for the measurement dynamics is non-trivial as the parameters depend on the current volume flow rate of the exhaust, which again depends on the current states in the exhaust pipe. Influences of the measurement dynamics are discussed in more detail in Sect. 3.6.2.

The validation of the  $\text{NO}_x$ , soot and torque model on the NEDC shows that the stationary model structure can be used to model the dynamic engine behaviour. Also transients, such as the gear shift at  $t = 980$  s, are well covered by the stationary models. Furthermore, the emission models react considerably faster to changes in emission formation than the measurements, as they do not suffer from measurement dynamics.

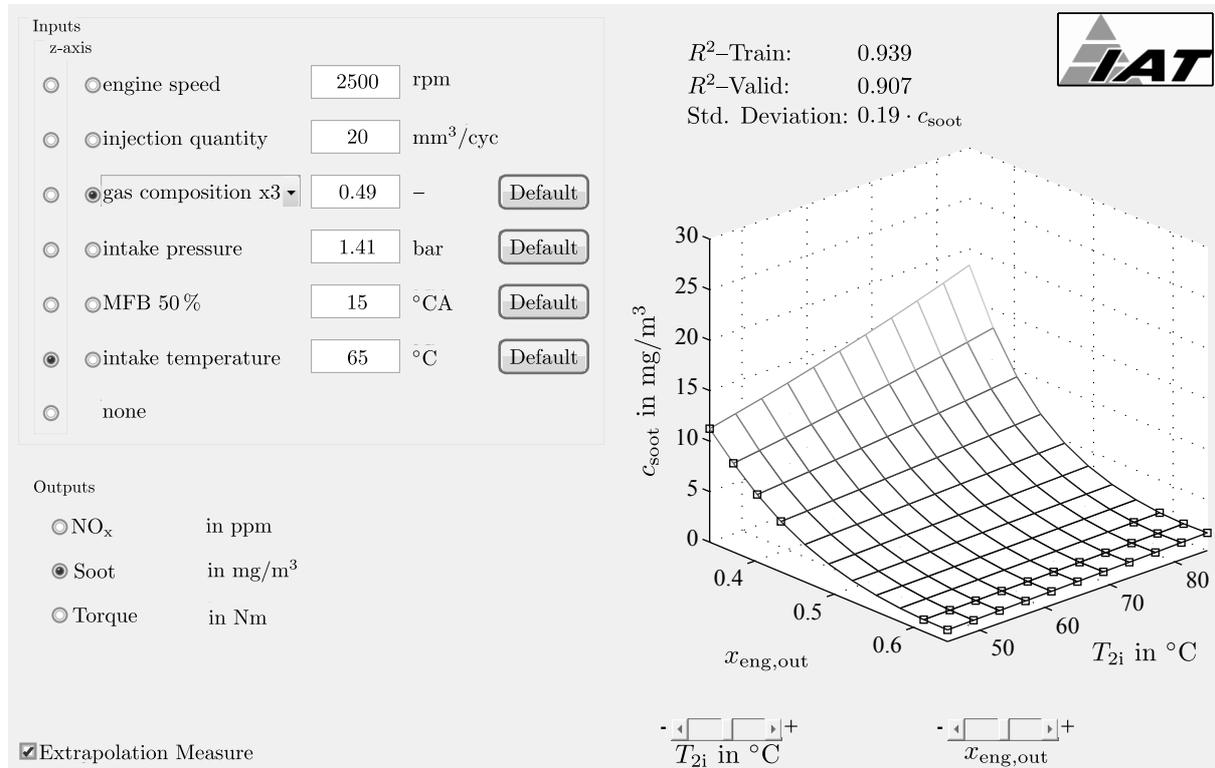
## 3.5 Analysis of the stationary emission characteristics

Since the presented experimental models can be regarded as black-box models, an analysis of the model characteristics based on the model parameters is only limitedly possible. Furthermore, as the models are high dimensional, a visualisation of the dependencies is non-trivial. In the following an *intersection plot tool* is introduced, which enables the visualisation and analysis of the dependencies with regard to one or two model inputs. After this, the  *$\text{NO}_x$ -soot trade-off* is introduced by such an intersection plot. This trade-off motivates the usage of model-based optimisation techniques to calibrate the engine control, as presented in Chap. 4.

### 3.5.1 Intersection plot tool

The presented models have a high dimensional input space. To enable an illustration of these high dimensional relations, an intersection plot tool is presented. For an intersection plot, all inputs except one is kept constant and the regarded output is plotted against the varying input. Thus, a two-dimensional plot arises that illustrates the dependency of the output on the selected input. If

two inputs are varied, the resulting output is depicted in a three-dimensional plot. The graphical user interface of the programmed intersection plot tool is shown in Fig. 3.11.



**Figure 3.11:** Intersection plot tool for the combustion models showing the soot emissions with regard to the gas composition after combustion  $x_{\text{eng,out}}$  and the intake temperature  $T_{2i}$ . Black squares in the plot show the extrapolation region of the model. Inputs are selected by the button group at the top left and the output is selected by the button group at the lower left. Sliders below the plot can be used for zooming.

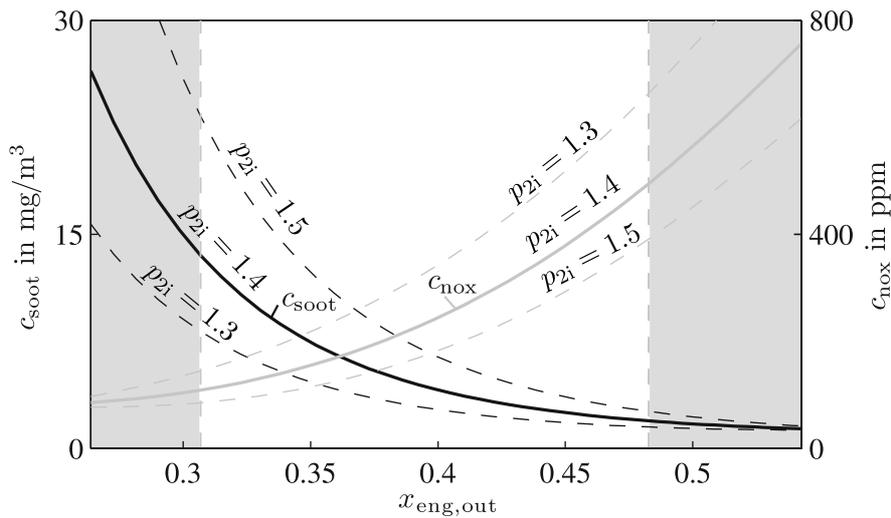
The plot shows the dependency of the soot emissions  $c_{\text{soot}}$  on the gas composition after combustion  $x_{\text{eng,out}}$  and the intake temperature  $T_{2i}$ . The selection of the inputs as varying variables can be seen by the button groups on the left of the plot. The black squares on the surface in the plot indicate the extrapolation region of the model. The calculation of this extrapolation measure, see eq. (3.21), can be disabled by the checkbox given in the lower left corner to speed up the intersection plot tool. The calculation of the extrapolation measure requires the determination of a convex hull and can be time consuming for higher dimensional relations. The determination of the here regarded four-dimensional convex hull requires only few milliseconds and is therefore not critical. The selection of the regarded model output is given by the button group below the inputs. Furthermore, two sliders are given underneath the plot for zooming the selected model inputs.

Given such a tool, the strong dependency of the soot emissions on the gas composition after combustion can be visualised. Furthermore, the influence of the intake temperature is shown. Besides an analysis of the dependencies, the tool can be applied to validate high dimensional models. The usage of the tool is in general independent of the applied model structure. Furthermore, it can be

adapted to alternative model inputs or can easily be changed to show the dependencies for other outputs. It is well suited to investigate the sensitivity of several inputs on the regarded output. This sensitivity is in the following investigated simultaneously for the  $\text{NO}_x$  and soot emissions.

### 3.5.2 $\text{NO}_x$ -soot trade-off

Using the intersection plot tool enables an investigation of the  $\text{NO}_x$ -soot trade-off. Figure 3.12 shows the intersection plots for  $\text{NO}_x$  and soot over the gas composition after combustion  $x_{\text{eng,out}}$ . It can be seen that  $x_{\text{eng,out}}$  has a crucial influence on the soot emissions, but also a strong influence on the  $\text{NO}_x$  emissions. These influence are in opposite directions, which is why there is a trade-off between those emissions. The dashed lines show the trade-off for different intake pressures. An increase of the pressure lowers the  $\text{NO}_x$  emissions, but increases at the same time the soot emissions. Similar dependencies can be shown for the other inputs. The shaded areas indicate the extrapolation regions of the models. It can be seen that the models are able to predict reasonable outputs in these regions.



**Figure 3.12:** Intersection plot of the simulated soot (black) and  $\text{NO}_x$  emissions (grey) depending on the gas composition after combustion  $x_{\text{eng,out}}$  for different intake pressures  $p_{2i}$ . All other inputs are fixed as  $n_{\text{eng}} = 2200$  rpm,  $u_{\text{inj}} = 24$   $\text{mm}^3/\text{cyc}$ ,  $T_{2i} = 70$   $^{\circ}\text{C}$  and  $\varphi_{\text{Q50}} = 14$   $^{\circ}\text{CA}$ .

The  $\text{NO}_x$ -soot trade-off is one of the main challenges in the calibration process of new engines. The legislative boundaries for both,  $\text{NO}_x$  and soot emissions, have become more stringent in recent years, see Fig. 1.1. Because of the high dimensional influences on the emissions, the complexity of the optimisation has increased simultaneously with the increased engine complexity. Hence, a calibration on the test bed is very time consuming and might not fully exploit the additional degrees of freedom. Therefore, a model-based optimisation is presented in Chap. 4.

## 3.6 Analysis of the dynamic emission characteristics

To analyse the dynamic emission characteristics, three typical transients are investigated in the following. Thereafter, the measurement dynamics are analysed in more detail.

### 3.6.1 Step in injection quantity and ramp in engine speed

In Fig. 3.13 three typical transients are presented. These are a ramp up and down in engine speed  $n_{\text{eng}}$ , a step up and down in injection quantity  $u_{\text{inj}}$  and a step up and down in injection quantity followed by a ramp up respectively down in engine speed. The latter can be regarded as acceleration respectively deceleration event. The air path states and the crank angle of main injection are adjusted by feedforward controls of their actuators. The desired values for  $m_{\text{air}}$  and  $p_{2i}$ , marked as dashed lines in the plot, are taken from a series calibration for stationary conditions. For the engine speed, a ramp is applied rather than a step, since it can not be changed instantaneously. On the other hand, the injection quantity can be adjusted cyclewise, why a step is applied.

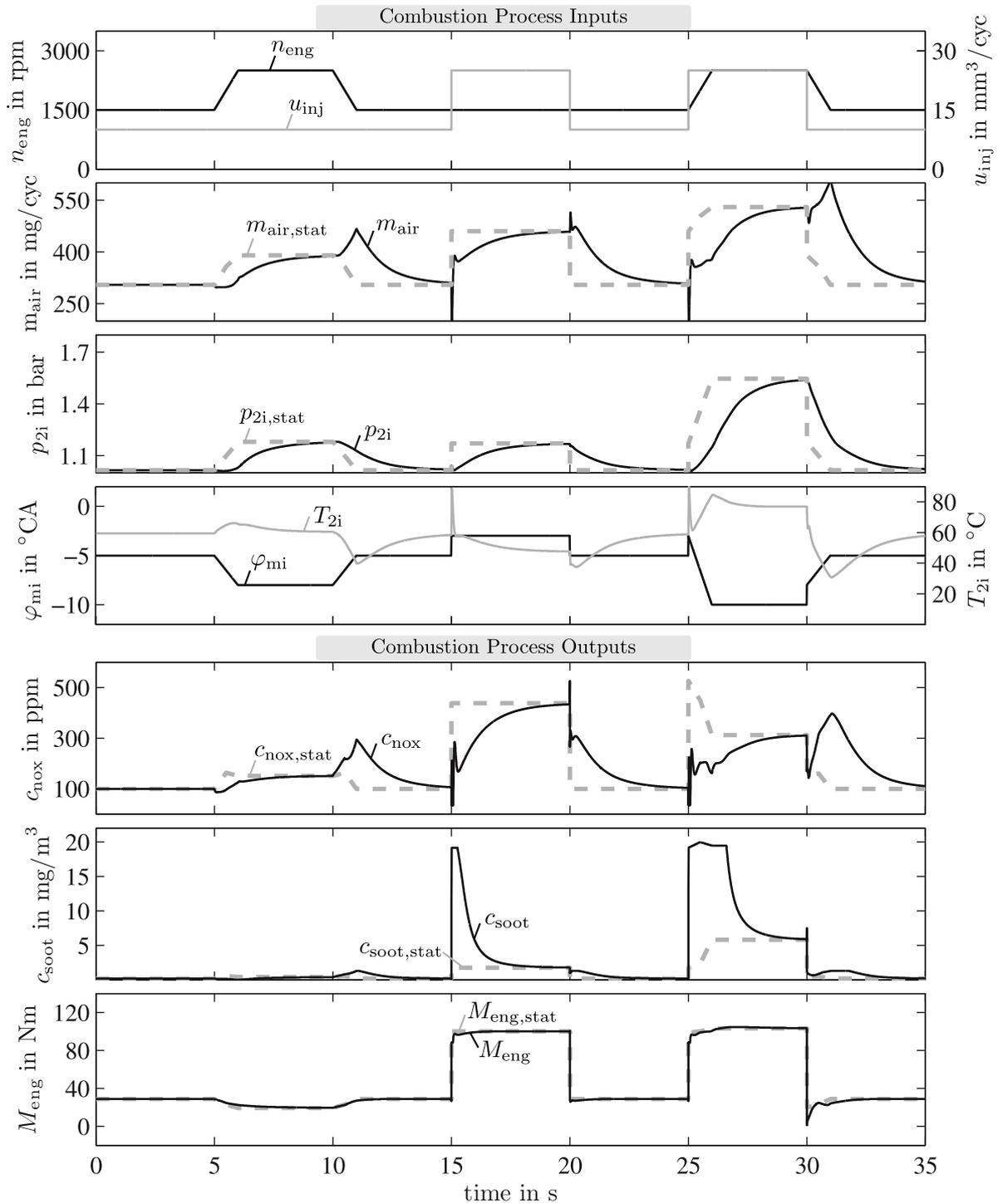
The presented results are derived from a simulation of the overall engine model, as introduced in Fig. 3.1. Therefore, the dynamic air path model, as presented in App. A, is applied together with the combustion model, presented in the previous sections. For the injection system no model is applied, since it is assumed that the injection characteristics follow their desired values without any dynamics.

At  $t = 5$  s a ramp in engine speed is shown. The desired values for  $m_{\text{air}}$  and  $p_{2i}$  are changed together with the change in engine operation point, see dashed lines in the plots. The air path states follow with significant dynamics, since they are adjusted by open loop controls of their actuators. The  $\text{NO}_x$  and soot emissions are marginally affected by the ramp up in engine speed at  $t = 5$  s, but the  $\text{NO}_x$  emissions increase for the ramp down at  $t = 10$  s. The ramp down in engine speed increases the air mass per cycle, which again leads to the formation of  $\text{NO}_x$ , which is negative.

For the step in injection quantity at  $t = 15$  s, the soot emissions are strongly affected. The stationary emissions are increased for both, the  $\text{NO}_x$  and the soot emissions, as is apparent by the dashed lines. The soot emissions overshoot their stationary value, whereas the  $\text{NO}_x$  emissions approach their stationary value from the lower side. Hence, the  $\text{NO}_x$  soot trade-off is also apparent for transient operations. For the step down, the  $\text{NO}_x$  emissions are higher than their stationary value, since the change in engine operation point also causes a change in desired values for  $m_{\text{air}}$  and  $p_{2i}$ .

The acceleration and deceleration events,  $t = 25$  s and  $t = 30$  s, show a similar behaviour as the separate investigated steps in injection quantity and ramps in engine speed, but the reactions of the emissions are more pronounced. This is due to a more significant change in desired values for  $m_{\text{air}}$  and  $p_{2i}$ . The acceleration event shows an overshoot in soot emissions, whereas the deceleration event shows an overshoot in  $\text{NO}_x$  emissions, which is due to the increased air mass per cycle.

Summarising, a step up in injection quantity poorly affects the soot emissions and a ramp down in engine speed poorly affects the  $\text{NO}_x$  emissions. The increased soot emissions are due to the



**Figure 3.13:** Dynamic simulation of the overall engine model, applying models for the dynamic air path and the stationary combustion process. Engine operation point trajectory is shown in the topmost plot. The first two transients illustrate separately the dynamics in engine speed and injection quantity, whereas the latter can be regarded as typical acceleration respectively deceleration event. The model actuators are adjusted by open loop controls, which is why there are relatively long settling times for  $m_{\text{air}}$  and  $p_{2i}$  (second and third plot). Engine outputs are shown in the lower three plots.

instantaneous increase of fuel mass and the consequential deficiency of air. The soot emissions are even more significant, if the step is followed by a ramp in engine speed, as is the case for the acceleration event. To lower the emissions for such transients, closed loop controls of the air path states are applied. Thus, the stationary states are reached faster and the overshoots are lowered. In Chap. 5 a non-linear offline optimisation of an acceleration event is regarded and closed loop controls are discussed. To further avoid high soot emissions during accelerations, a smoke limitation is presented in Sect. 5.4. On the other hand, the ramp down in engine speed poorly affects the  $\text{NO}_x$  emissions, because of the variations in air mass per cycle. The variations in air mass per cycle result from a change in the desired value for  $m_{\text{air}}$ , but also from the increased number of cycles per minute, determined by the engine speed.

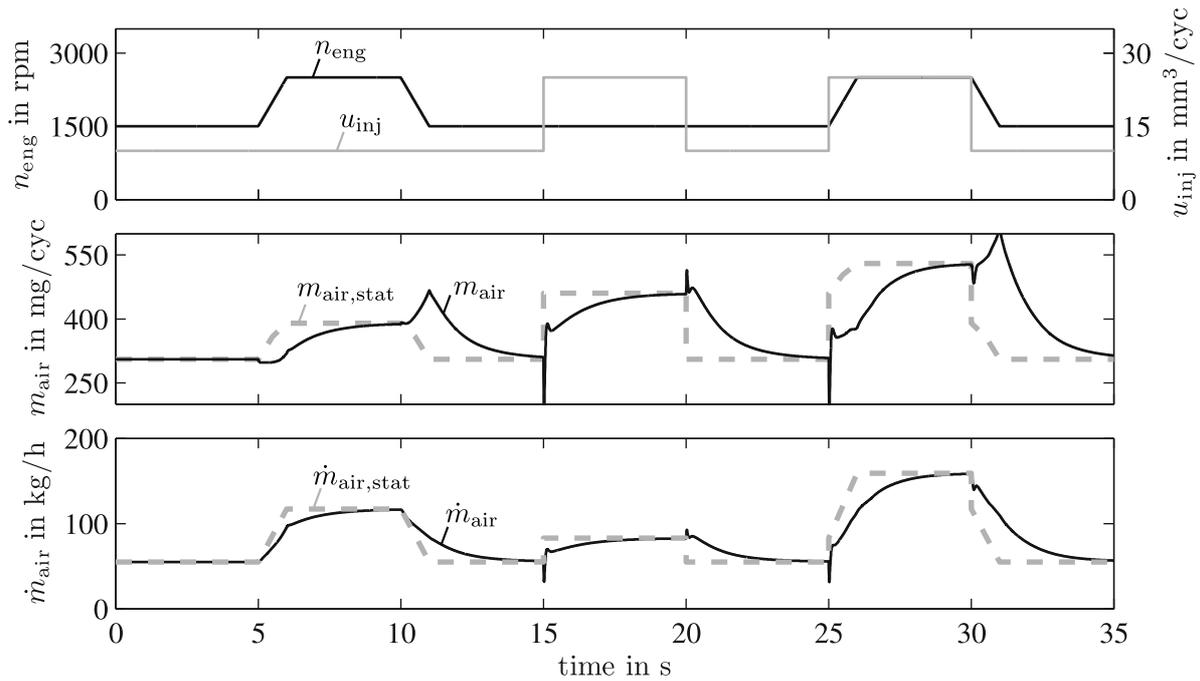
To analyse the increased  $\text{NO}_x$  emissions for a ramp down in engine speed, the air mass per cycle  $m_{\text{air}}$  and the air mass flow rate  $\dot{m}_{\text{air}}$  are regarded in Fig. 3.14. For the ramp down in engine speed at  $t = 10$  s and  $t = 30$  s, the air mass per cycle increases for the first second instead of decreasing, as is intended by its desired value. This is due to the influence of the engine speed on the air mass per cycle. Regarding the air mass flow rate,  $\dot{m}_{\text{air}}$  in the bottom plot in Fig. 3.14, it decreases from the beginning of the ramp, but the simultaneous decrease in engine speed (topmost plot) causes a higher air mass per cycle (middle plot) for the duration of the engine ramp,  $t = 10 - 11$  s and  $t = 30 - 31$  s. Since the combustion depends on the air mass per cycle rather than the air mass flow rate, this favours the formation of  $\text{NO}_x$ , as is observable in Fig. 3.13. In Sect.5.3 a coupled actuator control structure is introduced to reduce the dynamic deviations for these ramps down in engine speed..

The properties observed in this overall simulation are also observable for measurements from the engine test bed, as is shown in App. B.4. These measurements further validate the overall engine simulation with the connected models, as introduced in Fig. 3.1. A detailed investigation of the measurement dynamics is made in the following section applying dynamic measurements.

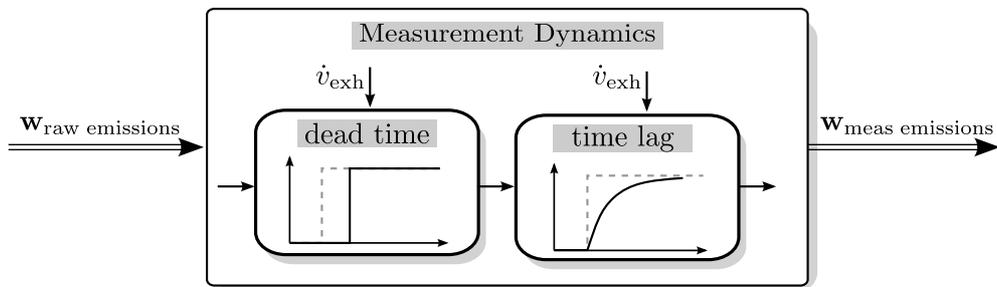
### 3.6.2 Influences of measurement dynamics

The measurement dynamics are modelled as dead time followed by a time lag, see Fig. 3.15 and the functional relation in eq. (3.1) with parameters given in Tab. 3.1. The parameters depend on the volume flow rate of the exhaust  $\dot{v}_{\text{exh}}$ , why the presented model is a time variant system. For such a time variant system the dead time block and the time lag block are not commutable, why the sequence of the two blocks is relevant.

The dead time is due to the travelling time of the emissions from their place of formation to the sensor. The time lag is due to mixing effects during this travelling time and due to sensor dynamics. The time lag due to mixing effects is therefore contributing simultaneously to the measurement dynamics as the dead time, why the separation of the dead time and the time lag block as in Fig. 3.15 is a simplification. To model the simultaneous influences during the travelling time, it is possible to split the dead time into a dead time before and after the time lag [3]. The time lag can then also be



**Figure 3.14:** Air mass per cycle in  $\text{mg}/\text{cyc}$  (middle plot) and air mass flow rate in  $\text{kg}/\text{h}$  (bottom plot) together with the engine operation point trajectory (topmost plot) for the transients presented in Fig. 3.13. The air mass flow rate approaches the desired value monotonously, whereas the air mass per cycle is influenced by the engine speed. It therefore changes in the wrong direction before approaching the desired value for ramps down in engine speed, see at  $t = 10 - 11$  s and at  $t = 30 - 31$  s.

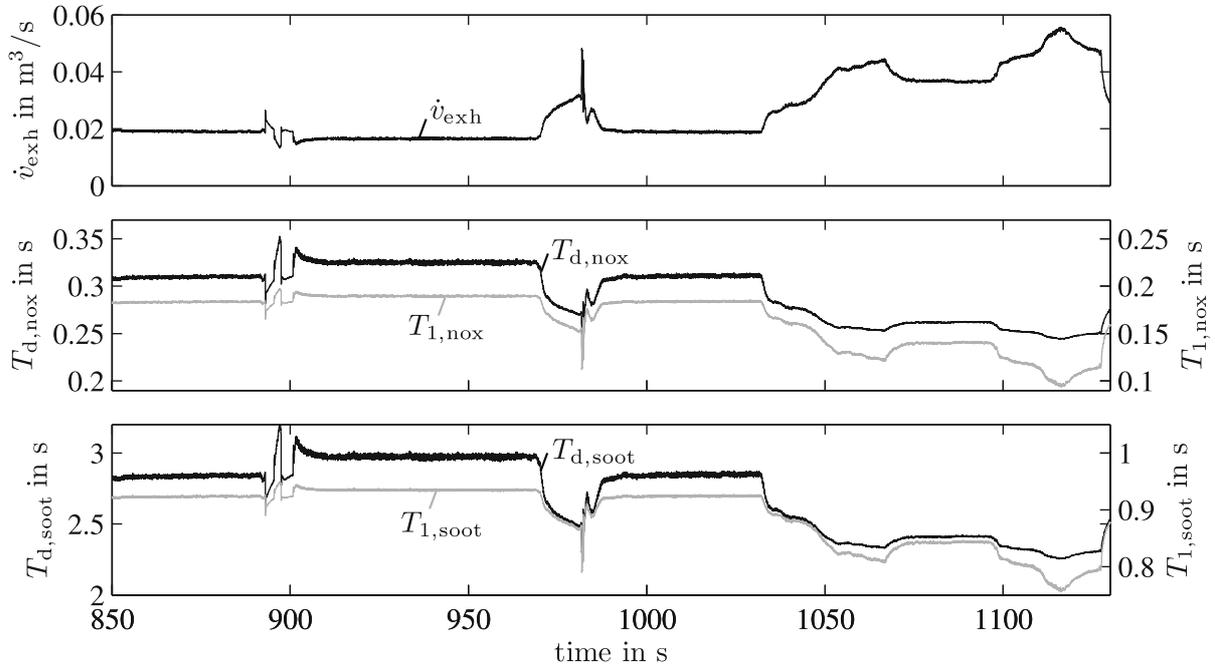


**Figure 3.15:** Block diagram of the modelled measurement dynamics. The dead time is followed by a time lag and both depend on the volume flow rate of the exhaust  $\dot{v}_{exh}$ . Thus, the produced raw emissions are delayed and smoothed before they are measured.

split into a time lag due to mixing effects and a time lag due to sensor dynamics. The differences are however relatively small, why the presented model is regarded to be sufficient for this application.

The variations in the dead times and the time lags are depicted for the extra urban part of the NEDC in Fig. 3.16 for the measurement dynamics of  $\text{NO}_x$  and soot. The variations for the  $\text{NO}_x$  emissions are less significant than for the soot emissions, which is due to the significant longer travelling distance for the soot measurement. The  $\text{NO}_x$  sensor is close to the outlet valve, whereas the soot

sensor extracts the exhaust gas by a separate hose. The dead times vary for the  $\text{NO}_x$  measurement from 0.25 s to 0.35 s and for the soot measurement from 2.2 s to 3.2 s. The time lags for  $\text{NO}_x$  are between 0.1 s and 0.2 s and for soot between 0.7 s and 1.0 s. Besides these variant time lags, there is a second time constant identified for the  $\text{NO}_x$  measurement at  $T_{2,\text{nox}} = 0.45$  s, see Tab. 3.1.

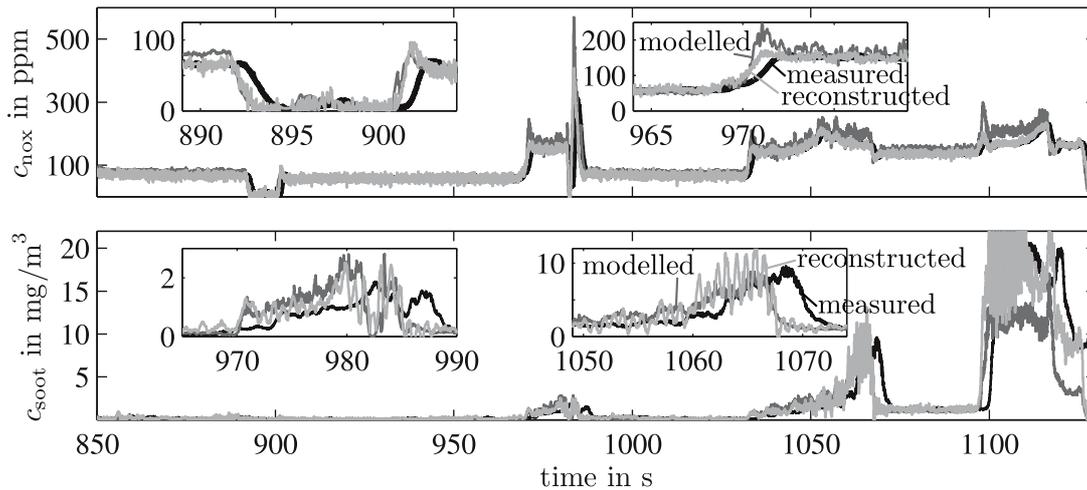


**Figure 3.16:** Dead times and time lags of the measurement dynamics for  $\text{NO}_x$  ( $T_{d,\text{nox}}$  and  $T_{1,\text{nox}}$  in the middle plot) and for soot ( $T_{d,\text{soot}}$  and  $T_{1,\text{soot}}$  in the lower plot) as a function of the volume flow rate of the exhaust ( $\dot{v}_{\text{exh}}$  in the topmost plot), shown for the extra urban part of the NEDC.

The variations in parameters depend on the volume flow rate of the exhaust  $\dot{v}_{\text{exh}}$ , shown in the topmost plot of Fig. 3.16. The volume flow rate is highly influenced by the engine speed, why the highest variations in parameters occur where the engine speed varies, which is the gear shift at  $t = 980$  s. These fast variations can stretch or shrink an emission peak if a change in emission formation is simultaneous with a change in volume flow rate of the exhaust. Hence, a true emission peak of the engine might be shorter or wider than the measured emission peak.

To reconstruct the produced raw emissions from the measured emissions, the measurement dynamics can be inverted. Due to the dependence of the parameters on the volume flow rate, the inversion is a non-trivial task. Problematic is the differentiating behaviour of the inverted time lag, which increases the measurement noise. The reconstructed emission signals are therefore filtered after inversion by a low pass filter to damp the high frequencies. The inverted and filtered signals are shown by the light grey lines in Fig. 3.17 together with the modelled (dark grey lines) and measured (black lines) emissions.

The stationary behaviour fits well, as is also shown previously in Fig. 3.10. Comparisons of the dynamics of the modelled raw emission signals (dark grey line) and the reconstructed signals (light



**Figure 3.17:** Measured, modelled and reconstructed  $\text{NO}_x$  and soot emissions for the extra urban part of the NEDC. The measured emission signals are shown by the black lines, the modelled raw emissions by the dark grey lines and the reconstructed raw emissions derived from an inversion of the measurement dynamics by the light grey lines.

grey line) are given by the zoomed plots. It shows that the dead times are well estimated. The inversion of the time lag increases the measurement noise, which is especially observable for the soot emissions. There is a middle-frequent oscillation on the measurement (time period  $\approx 1$  s), which is gained by the inversion and not filtered by the low pass filter. This oscillation is probably due to a closed loop control in the soot sensor and might be eliminated by a band stop filter. Besides these middle-frequent oscillations, the dynamics of the reconstructed signals fit well to the modelled raw emissions and with the applied low pass filter, the increase in measurement noise is acceptable for the  $\text{NO}_x$  emissions.

This shows, that the identified combustion models with the simulation in Fig. 3.10 show a good agreement with the measured behaviour, also for dynamic measurements.

### 3.7 Summary

Mean value models for the combustion outputs  $\text{NO}_x$ , soot and the engine torque are presented for a Diesel engine. The presented combustion models are part of an overall engine model, consisting of a model for the air path, a model for the injection system and the presented combustion models. The main dynamics of such a system result from the air path, why a dynamic model structure is required for the air path. The dynamics of the injection system are considerably faster. Injection characteristics, such as the crank angles of pilot and main injections, can be adjusted cyclewise without any dynamic. The dynamics of the rail pressure are neglected but can be included for future work. Hence, the states of the injection system are assumed to follow their desired values without any dynamic, why no model is required for the injection system.

For the combustion process an experimental model is identified. For the applied mean value model, the outcome of the combustion process depends only on the initial states and the boundary conditions of combustion. It is therefore modelled by a stationary model structure, which is trained by stationary measurements. Measurements dynamics, which are especially significant for emission measurements, are therefore avoided. The dynamics in emission formation are covered by the fast measured inputs of the combustion models or, if an overall engine simulation is applied, by the dynamic air path model.

The combustion process is modelled by a global-local model structure, applying local adaptive polynomials as local models for fixed engine operation points. The local models are distributed over the engine operation points such that the relevant region of test cycles is covered. The global model output is calculated by a bilinear interpolation of the local models over the engine operation points. For the local models, various model inputs are discussed. The suitability of an input depends on the modelled output and the regarded application. The applied local model inputs are the air mass per cycle  $m_{\text{air}}$ , the intake pressure  $p_{2i}$ , the crank angle of 50 % mass fraction burnt  $\varphi_{Q50}$  and the intake temperature  $T_{2i}$ . For soot, the gas composition after combustion  $x_{\text{eng,out}}$  is applied instead of  $m_{\text{air}}$ . Regarding the measured emission concentrations, transformations to the more meaningful emission mass flow rates are presented. Furthermore, a logarithmic transformation is shown for the soot emissions, which enables a more accurate identification of low emission values. This is advantageous for an optimisation, since there the accuracy for low emission values is more important. The engine torque is modelled without any transformation of the measured value.

The qualities of the  $\text{NO}_x$ , soot and torque model are stated for stationary training and validation data. The dynamic performance is shown for dynamic measurements and simulations of the NEDC. To investigate the characteristics of the stationary emission formation, an intersection plot tool is introduced that enables the visualisation of high dimensional influences. For an analysis of the dynamic emission formation, typical transients, such as an acceleration and a deceleration event, are regarded. Therefore, the overall engine model is simulated from the engine actuators to the raw emissions and the engine torque by connecting the dynamic air path model to the stationary combustion model. Furthermore, measurement dynamics are discussed and an inversion of those is presented.

---

## 4 Optimisation of the Stationary Engine Operation

---

In the following, the model-based optimisation of the stationary engine operation is presented for a local and a global approach. The local optimisation regards each engine operation point separately and several engine operation points are considered simultaneously for the global optimisation. Objective functions for both optimisations are introduced. Results of the local optimisation show the Pareto front of the  $\text{NO}_x$ -soot trade-off. For the global optimisation a robustness constraint is formulated such that results are non-sensitive to model uncertainties and series variations. Comparisons between the local optimisation results, the global optimisation results and the results from the robust global optimisation are given and a dynamic simulation of the NEDC is shown.

---

Due to the increased complexity of combustion engines, the optimisation of the engine control is a major task of the engine development process. This high dimensional non-linear problem can hardly longer be solved on the engine test bed alone. To fully utilise the potentials of modern combustion engines, model-based approaches have to be applied. The goal of the optimisation is the minimisation of the fuel consumption, which is demanded by the customer and by the legislation. At the same time, emission limits need to be fulfilled, which might be in contradiction to the reduction of fuel consumption. Major challenges in the fulfilment of emission limits are, especially for the Diesel engine, the  $\text{NO}_x$  and soot emissions, which behave in opposite directions.

The optimisation of the engine control is divided into a stationary optimisation and a dynamic optimisation. The stationary optimisation provides set points for the air path states and the injection system. During engine operation, these set points are adjusted by the engine actuators, which are controlled by open and closed loop controls. The design of these controls is the objective of the dynamic optimisation, which is addressed to in the following chapter.

A stationary optimisation assumes that all dynamic effects are decayed. Hence, only the combustion models are applied without the connection to the dynamic air path model. For the local optimisation, one engine operation point is regarded, whereas several operation points are regarded simultaneously for the global optimisation. The relevance of an operation point is determined by the test cycle over which the legislative boundaries and the fuel consumption are defined. Since the global optimisation is a high dimensional problem, good initial values are required. These initial values are provided by the local optimisation, which is performed prior to the global optimisation.

The applied combustion models are models for the engine torque, the  $\text{NO}_x$  and the soot emissions as presented in Chap. 3. The utilised model inputs are the variables to be optimised. The number of model inputs should therefore be identical to the number of engine actuators such that the op-

timised set points correspond to a unique actuator setting. If the number of model inputs exceeds the number of actuators, some inputs need to be regarded as constant parameters. The set points are then optimised and stored according to these parameters. This is for example the case, if the influences of ambient temperature or pressure are regarded. These ambient conditions can not be adjusted by actuators, but an inclusion of these might be necessary for an accurate model. The ambient conditions are then regarded as constants for an optimisation and the optimised settings are stored according to the assumed ambient conditions. On the other hand, if the number of actuators exceeds the number of model inputs, their setting is not unique as one or more degrees of freedom are unused. Then one or more actuators need to be hold constant or be coupled to other actuators to reduce the degrees of freedom.

A central role in the optimisation process is the formulation of an objective function. Possible objective functions for local and global optimisations are discussed in Sect. 4.1. Thereafter, local and global optimisation results are presented in Sect. 4.2 and Sect. 4.3 respectively. Model uncertainties and series variations are regarded. Finally the optimisation results are discussed in Sect. 4.4. Presented results are given for an 1.9 l Common-Rail Diesel engine, certified for Euro 4 emission standard. The goal is to optimise this engine with regard to the Euro 5 emission standard.

## 4.1 Objective function

To derive an objective function, the goal of the optimisation and the legislative boundaries need to be translated into a functional expression. The goal of the optimisation is to minimise the fuel consumption and the regarded legislative boundaries are the limit values for the  $\text{NO}_x$  and soot emissions. Since the emission masses are limited, the modelled emission mass flow rates are applied rather than the measured emission concentrations. An optimisation of the emission concentrations is not equivalent, as the emission mass depends also on the volume flow rate of the exhaust. Thus, a high emission concentration might be superior to a low concentration if the volume flow rate of the exhaust is significantly smaller.

Besides the optimisation with regard to an objective function, it needs to be guaranteed that the obtained results are reachable in practice. This is ensured by utilising experimental models and by checking in every optimisation step, if the regarded settings are in the drivability space of the applied model.

### 4.1.1 Criterion for a local optimisation

For the local optimisation, only one operation point is regarded. Hence, the fuel consumption is given since it is defined by the regarded operation point, which again is defined by  $u_{inj}$  and  $n_{eng}$  (and e. g. not by the engine torque  $M_{eng}$ ). Furthermore, the emission limits are formulated over a test cycle with varying operation points, why there are no emission constraints given for a local optimisation. It is therefore not possible to translate the goal of the optimisation into an objective

function for the local optimisation, which is why the objective function is chosen as a weighted multi-objective criterion. Such an objective function can then be adjusted by its weights. Since the emissions  $\text{NO}_x$  and soot are major challenges, the criterion is formulated with regard to these emissions,

$$\begin{aligned} \min_{\mathbf{u}} J_{\text{local}}(\mathbf{u}) &= k_{\text{nox}} \dot{m}_{\text{nox}}(\mathbf{u}) + k_{\text{soot}} \dot{m}_{\text{soot}}(\mathbf{u}) \\ \text{for } \{n_{\text{eng}}, u_{\text{inj}}\} &= \text{const.} \end{aligned} \quad (4.1)$$

The variable  $\mathbf{u}$  is four-dimensional and contains the local model inputs

$$\mathbf{u}^T = [m_{\text{air}}, p_{2i}, \varphi_{Q50}, T_{2i}]. \quad (4.2)$$

The weights  $k_{\text{nox}}$  and  $k_{\text{soot}}$  are chosen such that

$$\begin{aligned} k_{\text{nox}} &= 1 - k_{\text{soot}} \\ \text{with } \tilde{k}_{\text{nox}} &\in [0, 1]. \end{aligned} \quad (4.3)$$

Different weights can be applied to get various ratios of  $\text{NO}_x$  to soot emissions. Since the emission limits for  $\text{NO}_x$  and soot are of different orders of magnitude (180 mg/km of  $\text{NO}_x$  to 5 mg/km of soot for the Euro 5 emission standard), the weights are varied in non equidistant steps. In App. D.5 a distribution of weights is derived which results into

$$\begin{aligned} k_{\text{nox}} &= \frac{(36^{2\tilde{k}_{\text{nox}}})}{35 \cdot 37} \\ \text{with } k_{\text{nox}} &\text{ equidistantly distributed on } [0, 1]. \end{aligned} \quad (4.4)$$

Thus,  $\tilde{k}_{\text{nox}} = 0$  implies  $k_{\text{nox}} = 0$  and  $k_{\text{soot}} = 1$  such that only soot is optimised,  $\tilde{k}_{\text{nox}} = 1$  implies  $k_{\text{nox}} = 1$  and  $k_{\text{soot}} = 0$  such that only  $\text{NO}_x$  is optimised and  $\tilde{k}_{\text{nox}} = 0.5$  implies the ratio 1 to 36 for  $k_{\text{nox}}$  to  $k_{\text{soot}}$  (corresponding to the ratio 5 to 180) such that the weighting of  $\text{NO}_x$  to soot equals the Euro 5 emission standard ratio. Optimising eq. (4.1) with these varying weights results in a *Pareto front*. For calibrated values on the Pareto front, none of the emissions can be lowered without increasing the other.

### Constraint

A constraint is formulated to guarantee that the obtained results are reachable in practice. Therefore, the drivability spaces  $\mathcal{D}_j$ , as defined in Sect. 3.3, are utilised. If the  $j$ -th local model is regarded for optimisation, the constraint

$$\mathbf{u} \in \mathcal{D}_j \quad (4.5)$$

needs to be checked in each optimisation step. This is a discontinuous constraint, as the regarded point  $\mathbf{u}$  can either be inside or outside the drivability space  $\mathcal{D}_j$ . Since continuous constraints are superior for most optimisation algorithms, the constraint is reformulated to

$$\text{Vol} \{\mathcal{D}_j \cup \mathbf{u}\} - \text{Vol} \{\mathcal{D}_j\} = 0. \quad (4.6)$$

This constraint also requires that a point  $\mathbf{u}$  is inside the drivability space, since the volume of the drivability space plus the input  $\mathbf{u}$ ,  $\text{Vol}\{\mathcal{D}_j \cup \mathbf{u}\}$ , is the same as the volume of the drivability space alone,  $\text{Vol}\{\mathcal{D}_j\}$ , if and only if  $\mathbf{u} \in \mathcal{D}_j$ . The advantage of such an formulation is however that an information about the distance of the point  $\mathbf{u}$  to the drivability space  $\mathcal{D}_j$  is given, since the volume  $\text{Vol}\{\mathcal{D}_j \cup \mathbf{u}\}$  increases with the distance of  $\mathbf{u}$  to  $\mathcal{D}_j$ . The formulation of a non-linear constraint as in eq. (4.5) or eq. (4.6) complicates the optimisation process, but it is necessary to ensure that the obtained results are reachable in practice.

### Alternative objective functions

It is also possible to include further terms, like the engine torque  $M_{\text{eng}}$ , in the multi-objective criterion of eq. (4.1) by applying additional weights

$$\min_{\mathbf{u}} J_{\text{local}}(\mathbf{u}) = k_{\text{nox}} \dot{m}_{\text{nox}}(\mathbf{u}) + k_{\text{soot}} \dot{m}_{\text{soot}}(\mathbf{u}) - k_{\text{torque}} M_{\text{eng}}(\mathbf{u}). \quad (4.7)$$

Drawbacks of such a criterion are the heuristic choices of the weights and the correlation between  $\text{NO}_x$  and torque. A high torque correlates to high  $\text{NO}_x$  emissions, why such an objective function is insensitive to variations between the weights  $k_{\text{nox}}$  and  $k_{\text{soot}}$ .

Another alternative objective function can be determined by defining an average emission limit for the emission mass flow rate

$$\bar{\dot{m}}_{\text{nox,limit}} = \frac{s_{\text{cycle}} \text{NO}_{x,\text{limit}}}{t_{\text{cycle}}}, \quad (4.8)$$

where  $s_{\text{cycle}}$  is the covered distance of the test cycle in km,  $\text{NO}_{x,\text{limit}}$  is the limit value for  $\text{NO}_x$  in mg/km and  $t_{\text{cycle}}$  is the required time for the test cycle.  $\bar{\dot{m}}_{\text{soot,limit}}$  can be defined analogue. Instead of minimising the fuel consumption, the local torque can then be maximised with regard to these limits,

$$\max_{\mathbf{u}} J_{\text{local}}(\mathbf{u}) = M_{\text{eng}}(\mathbf{u}) \quad (4.9)$$

with

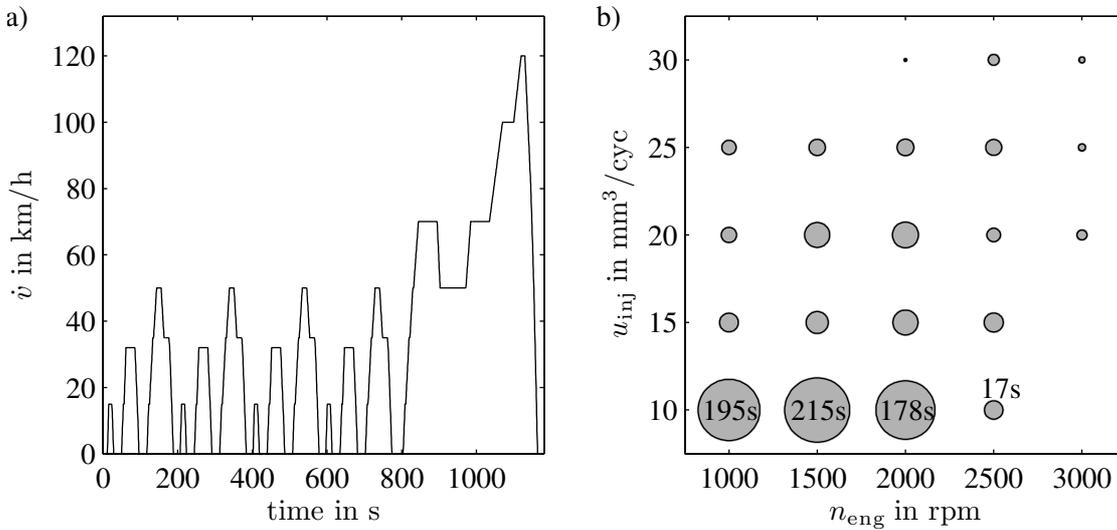
$$\dot{m}_{\text{nox}}(\mathbf{u}) \leq \bar{\dot{m}}_{\text{nox,limit}} \quad (4.10a)$$

$$\dot{m}_{\text{soot}}(\mathbf{u}) \leq \bar{\dot{m}}_{\text{soot,limit}}. \quad (4.10b)$$

Such an optimisation assumes an equal distribution of the produced emissions on the local operation points. This is however unfavourable, as operation points with high engine speeds and injection quantities produce relatively high emissions compared to operation points with low engine speeds and low injection quantities. Therefore and because of the easy interpretation and good visualisation of the  $\text{NO}_x$ -soot trade-off, the multi-objective criterion as in eq. (4.1) is applied for the local optimisation, with the non-linear constraint as stated in eq. (4.6). The engine torque is then regarded for the global optimisation, for which the optimisation criterion is discussed in the following.

### 4.1.2 Criterion for a global optimisation

For the global optimisation, all  $M$  operating points are regarded simultaneously. It is therefore necessary to relate these operation points to each other, for which the local weights  $k_j$  are applied. Since the fuel consumption and the emission limits are defined over a test cycle, this test cycle is applied to derive these weights. Here the NEDC is regarded, which is depicted in Fig. 4.1 a). The local weights are derived from the frequency of occurrence of an operation point in the test cycle, given as summed up residence time in s. The individual weights are indicated by the sizes of the circles in Fig. 4.1 b). Operation points with low engine speeds and low injection quantities dominate the NEDC. Note that the optimisation can be performed with regard to alternative emission standards with different test cycles by changing the weights accordingly.



**Figure 4.1:** a) Velocity diagram of the NEDC and b) derived weights  $k_j$  for the local engine operation points. The sizes of the circles indicate the magnitudes of the weights in s. Weights are derived from the frequency of occurrence of an engine operation point in the NEDC and they are utilised to relate the operation points to each other for the objective function of the global optimisation. The weights for the first four operating points are stated in the plot.

The goal of the engine optimisation is to minimise the *brake specific fuel consumption (bsfc)* over the regarded test cycle, while fulfilling the legislative boundary conditions. The bsfc is defined as summed up fuel consumption over the produced engine work. The optimisation goal can then be translated into

$$\min_{\mathbf{u}} J_{\text{global}}(\mathbf{u}) = \text{bsfc}(\mathbf{u}) \quad (4.11a)$$

$$= \frac{\sum_{j=1}^M k_j \dot{m}_{f,j}}{\sum_{j=1}^M k_j W_j(\mathbf{u})} \quad \text{in } \frac{\text{g}}{\text{kWh}} \quad (4.11b)$$

$$= \frac{\sum_{j=1}^M k_j \frac{n_{\text{eng},j}}{60} \rho_{\text{Diesel}} u_{\text{inj},j} \frac{4}{2}}{\sum_{j=1}^M k_j 2\pi \frac{n_{\text{eng},j}}{60} M_{\text{eng},j}(\mathbf{u})}, \quad (4.11c)$$

with the constraints for the Euro 5 emission standard

$$\sum_{j=1}^M \frac{k_j \dot{m}_{\text{nox},j}(\mathbf{u})}{s_{\text{cycle}}} \leq \text{NO}_{x,\text{limit}} \quad \text{in } \frac{\text{mg}}{\text{km}} \quad (4.12a)$$

$$\sum_{j=1}^M \frac{k_j \dot{m}_{\text{soot},j}(\mathbf{u})}{s_{\text{cycle}}} \leq \text{soot}_{\text{limit}} \quad \text{in } \frac{\text{mg}}{\text{km}}, \quad (4.12b)$$

and the constraint to guarantee reachable results, similar to eq. (4.6),

$$\mathbf{u} \in (\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_M). \quad (4.13)$$

The bsfc is given in g/kWh. The numerator describes the injected fuel mass over the test cycle and the denominator the produced energy. The division by two in eq. (4.11c) is because a four stroke engine has one injection every two turns for each cylinder and the multiplication by four is because a four cylinder engine is regarded. The accumulated emissions over the test cycle in eq. (4.12) are divided by the covered distance of the test cycle  $s_{\text{cycle}}$ , since the limiting values  $\text{NO}_{x,\text{limit}}$  and  $\text{soot}_{\text{limit}}$  are given in mg/km,  $\text{NO}_{x,\text{limit}} = 180 \text{ mg/km}$  and  $\text{soot}_{\text{limit}} = 5 \text{ mg/km}$ .

Since the engine speed  $n_{\text{eng},j}$  and the fuel injection quantity  $u_{\text{inj},j}$  are constant for a given engine operation point, only the engine torque  $M_{\text{eng},j}$  in eq. (4.11) depends on the local model inputs  $\mathbf{u}$ . Thus, the objective function in eq. (4.11) can be reformulated to maximise the torque weighted by the time periods  $k_j$ ,

$$\max_{\mathbf{u}} J_{\text{global}}(\mathbf{u}) = \sum_{j=1}^M k_j 2\pi \frac{n_{\text{eng},j}}{60} M_{\text{eng},j}(\mathbf{u}). \quad (4.14)$$

Hence, the objective of the global optimisation is the maximisation of a time period weighted sum of local powers, which again depend on the local torques  $M_{\text{eng},j}$ . The criterion in eq. (4.14) can therefore be regarded as the maximisation of the sum of local works. It shall be mentioned that a change in local torques also causes a change in the local weights  $k_j$  derived from the NEDC, since for a given vehicle setting the requested torque is determined by the regarded driving cycle. Depending on the produced local torques, the required injection quantities for these torque requests can vary. However, as these variations are relatively small and since a consideration would highly complicate the optimisation process, it is neglected here. Hence, constant weights  $k_j$  are assumed.

As all inputs  $\mathbf{u}$  of the objective function are optimised simultaneously, a high dimensional problem arises.  $M = 21$  local models are regarded and four local model inputs are optimised. Thus, the optimum has to be searched in 84 dimensions (degrees of freedom). It is therefore essential to have good initial values, why the local optimisation is performed prior to the global optimisation.

Alternative to the presented objective function, a weighted multi-objective function as in eq. (4.7) can also be utilised as objective function for the global optimisation. The weights of such a multi-objective criterion would then be adjusted by trial and error, such that the emission standards are fulfilled. The determination of suited weights is a laborious task and complicates the optimisation process, why the maximisation of eq. (4.14) is recommended and applied here.

## 4.2 Local optimisation

For the local optimisation, the weighted multi-objective function of eq. (4.1) is applied, with the drivability constraint as in eq. (4.6). Various weights are applied for optimisation, distributed as given by eq. (4.3) with eq. (4.4). The optimisation is performed in all  $M = 21$  operation points separately. There are various optimisation algorithms possible to determine the optimum. A brief overview of these algorithms is given in the following. Thereafter, the local optimisation results are presented and discussed.

### 4.2.1 Numerical optimisation algorithms

Because of the non-linear constraint in eq. (4.6), the optimisation problem can not be solved analytically. Therefore, numerical optimisation algorithms are applied to determine the minimum of the objective function. Optimisation algorithms can in general be divided into local and global optimisation algorithms. Such a classification in local and global regards only the numerical algorithm and in particular not whether a local or a global engine optimisation is performed.

Local optimisation algorithms start at an initial value and move in each optimisation step in the direction of the next local optimum. The various local optimisation algorithms can be distinguished by how the direction and the step length is determined. They further can be classified into constrained and unconstrained methods. The simplest methods calculate the search direction from several evaluations of the objective function. More complex methods require one or more derivatives of the objective function to determine the next optimisation step. The latter methods are computationally more expensive, but require in general less optimisation steps. A representative of the simple methods is the *Nelder-Mead simplex algorithm* [116] and of the more complex methods the *steepest descent algorithm* [31, 39]. Further local optimisation methods utilise knowledge about the form of the objective function. A representative of this class is the *Levenberg-Marquardt algorithm* [98], which assumes a quadratic objective function (non-linear least squares). A representative of the constrained optimisation algorithms is *sequential quadratic programming* [125]. Constraints can also be incorporated into an unconstrained algorithm by adding a penalty term to the objective function. Overviews of local optimisation algorithms are given in [39, 125].

If several local optima exist, it is possible that local optimisation algorithms get stuck in a local optimum and can not find the global optimum. Then, global optimisation algorithms need to be applied. These contain a stochastic element that enables them to leave a local minimum. Global optimisation algorithms follow predefined rules, which are based on physical or biological observations. Common algorithms are *simulated annealing* [79], *particle swarm optimisation* [76] and *genetic algorithms* [132, 153]. These algorithms can be adapted to the particular problem by several parameters, e. g. the cooling rate of the simulated annealing algorithm. It is also possible to combine local and global optimisation techniques, e. g. by determining the initial values randomly and performing for each initial value a local optimisation [153]. However, also global optimisation algorithms can not guarantee to find the global optimum.

### 4.2.2 Local optimisation results

There are several effects which influence the local optimisation results. These are the dependency on the utilised optimisation algorithm, the influence of the intake temperature and the deviations due to model uncertainties and series variations. These influences and the local optimisation results are presented in the following.

#### Dependency on the optimisation algorithm

For the local optimisation, either a local or a global optimisation algorithm can be utilised. A local optimisation algorithm requires to determine an initial value, which can be obtained from a-priori knowledge, such as the series settings of a preceding engine. If no a-priori information is given, the initial value shall at least be in the drivability space. The initial value is then determined as the centre of gravity of the regarded drivability space.

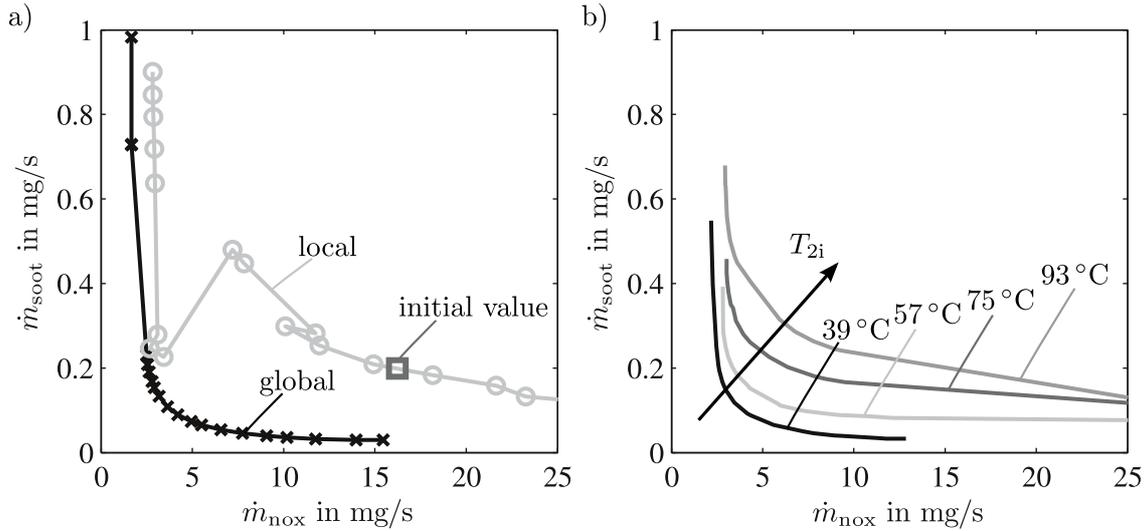
For the local optimisation algorithm, a sequential quadratic programming is utilised. The derivatives of the objective function are determined analytically, which is one of the advantages of the applied local polynomial models. For the global optimisation, a combination of simulated annealing, particle swarm optimisation and genetic algorithms is applied. The applied function is available in MATLAB and is described in detail in [126]. An adaption of the optimisation algorithms on the optimisation problem could further accelerate the optimisation process. The design of an optimisation algorithm is however not the task here.

Results determined by the local optimisation algorithm are shown by the grey line in Fig. 4.2 a). The various results, indicated as circles, are optimisation results with varying weights,  $k_{\text{nox}}$  and  $k_{\text{soot}}$ . For several weights, the local optimisation algorithm gets stuck in a local minimum close to the initial value, which shows the crucial influence of the initial value for a local optimisation algorithm. The results of the global optimisation algorithm are shown in Fig. 4.2 a) by the black line and the crosses. These are superior to the results of the local optimisation algorithm and show the Pareto front of the  $\text{NO}_x$ -soot trade-off for the regarded operation point. It can further be seen that the distribution of weights, as suggested in eq. (4.4), results in a good distribution of results in the relevant region of the Pareto front

The results in Fig. 4.2 a) show that the dependency on the initial value can be avoided if a global optimisation algorithm is utilised. Such an algorithm requires a longer computing time, but is more reliable in finding the global optimum. The local optimisation algorithm does however require good initial values, e. g. from a previous engine calibration.

#### Temperature dependency

The local optimisation of eq. (4.1) offers 4 optimisation variables, air mass per cycle  $m_{\text{air}}$ , intake pressure  $p_{2i}$ , crank angle of 50% mass fraction burnt  $\varphi_{Q50}$  and intake temperature  $T_{2i}$ , see eq. (4.2). The optimised settings of these variables need then to be controlled during engine operation via



**Figure 4.2:** a) Local optimisation results in dependence on the applied optimisation algorithm showing the engine operation point  $n_{\text{eng}} = 2000$  rpm and  $u_{\text{inj}} = 25$  mm<sup>3</sup>/cyc. Grey: optimisation results gained from a local optimisation algorithm (here: *sequential quadratic programming*) with the initial value as plotted; black: optimisation results gained from a global optimisation algorithm (here: a combination of *simulated annealing*, *particle swarm optimisation* and *genetic algorithms*). b) Local optimisation results in dependence on the intake temperature for the same operation point. For each Pareto front a fixed intake temperature is assumed for optimisation.

the engine actuators.  $m_{\text{air}}$  is mainly adjusted by the egr-valve  $s_{\text{egr}}$  and  $p_{2i}$  by the turbocharger guide vanes  $s_t$ , both with coupling effects on each other.  $\varphi_{Q50}$  is mainly controlled by the crank angle of main injection  $\varphi_{\text{mi}}$ , usually applying an open loop control. The intake temperature  $T_{2i}$  is more complicated to be controlled. It is influenced by the approaching cooling air mass flow on the intercooler, which however depends on the velocity of the vehicle and is therefore not adjustable. Besides this, the temperature is also influenced by the ratio of low-pressure egr to high-pressure egr. The low-pressure egr is double cooled and therefore significantly colder than the high-pressure egr. Thus, the proportion of low-pressure and high-pressure egr can be utilised as actuating variable, see [107]. On the other hand, if no low-pressure egr is available, the intake temperature can not be influenced.

Due to this limited influence, it might be necessary to regard the intake temperature as a constant parameter for the optimisation. The optimisation of eq. (4.1) reduces then to a three dimensional problem and needs to be solved with regard to several constant intake temperatures. The optimisation results are then stored subject to various constants. The optimisation result with various constant intake temperatures are shown in Fig. 4.2 b). The Pareto front is shifted with an increasing temperature towards higher emissions, without significant changes in shape. It shows a remarkable influence on the emissions. Hence, a low intake temperature is superior to a high intake temperature, why low-pressure egr shall be used whenever possible. Such information about the dependencies

of the emissions on the intake temperature can further be applied to determine the required size of an intercooler.

### Model uncertainties

Model uncertainties can lead to unreachable and overrated results. These uncertainties need therefore be regarded at an early state of the engine optimisation. The model uncertainties are determined by the simulation qualities of the local models in the region of the predicted optimum. Let  $\mathbf{u}_{\text{opt}}$  be the vector of the predicted optimum with the corresponding modelled output  $\hat{y}$ , then the model uncertainty at  $\mathbf{u}_{\text{opt}}$  is defined as

$$\Delta \dot{m}_{\text{mod}}(\mathbf{u}_{\text{opt}}) = \pm \sqrt{\text{Cov}(\hat{y})}. \quad (4.15)$$

The variance of the estimate  $\hat{y}$  can be calculated by

$$\text{Cov}(\hat{y}) = \text{Cov}(\mathbf{x}^T \hat{\mathbf{w}}) \quad (4.16a)$$

$$= \mathbf{x}^T \text{Cov}(\hat{\mathbf{w}}) \mathbf{x}. \quad (4.16b)$$

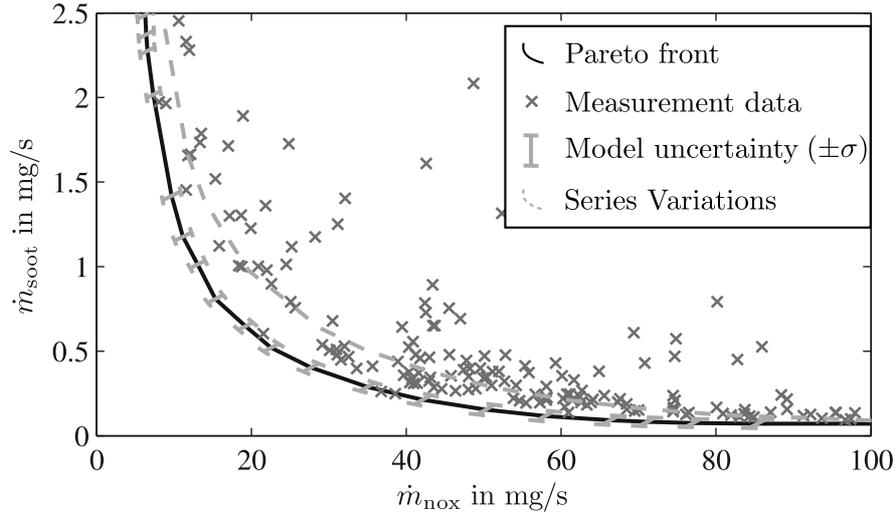
Note that the variance of  $\hat{y}$  is here denoted by its covariance, as for scalars holds that  $\text{Var}(\hat{y}) = \text{Cov}(\hat{y})$ . The covariance matrix of  $\hat{\mathbf{w}}$  is determined at model training. The variables  $\mathbf{x}$  are the regressors of the local polynomial model, see eq. (3.19), for the predicted optimum  $\mathbf{u}_{\text{opt}}$ . Since the parameters are estimated with a least squares algorithm, the covariance matrix is calculated with the input matrix  $\mathbf{X}$  and the estimated noise variance  $\hat{\sigma}^2$ , see eq. (2.37), by

$$\text{Cov}(\hat{\mathbf{w}}) = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}. \quad (4.17)$$

With the definition as given in eq. (4.15) and the relations as stated in eq. (4.16) and eq. (4.17), a model uncertainty can be defined for the  $\text{NO}_x$  model,  $\Delta \dot{m}_{\text{nox,mod}}(\mathbf{u}_{\text{opt}})$ , and the soot model,  $\Delta \dot{m}_{\text{soot,mod}}(\mathbf{u}_{\text{opt}})$ . The influences of the model uncertainties on the local optimisation results,  $\dot{m}_{\text{opt}} \pm \Delta \dot{m}_{\text{mod}}$ , are depicted in Fig. 4.3 by the grey confidence intervals on the Pareto front. As the variations are relatively small, the model gives reliable results. The measurement data is plotted as crosses in the plot and the crosses close to the Pareto front indicate that the found optima are reachable in practice. Since there are no crosses on the lower left side of the Pareto front, it can further be concluded that the found optima are global optima.

### Series variations

Optimisation results obtained for a single test engine are, due to series variations, not directly transferable to a series production. The influences of series variations are therefore considered by the derivatives of the emission models,  $\partial \dot{m}_{\text{nox}}(\mathbf{u})/\partial \mathbf{u}$  and  $\partial \dot{m}_{\text{soot}}(\mathbf{u})/\partial \mathbf{u}$ . Since the local models are described as polynomials, see Sect. 3.2, these derivatives can be determined analytically. The derivatives can be interpreted as sensitivities of the emissions on the model inputs. To regard the



**Figure 4.3:** Pareto front for the  $\text{NO}_x$  and soot emissions at the engine operation point  $n_{\text{eng}} = 2500$  rpm and  $u_{\text{inj}} = 30 \text{ mm}^3/\text{cyc}$ . Model uncertainties are indicated by the confidence intervals on the Pareto front. Measurement data, which is applied only for model training, is shown as crosses for a better visualisation of the Pareto front derived from the optimisation. Deviations due to worst case assumptions for series variations are plotted by the dashed grey line.

worst case influences of input variations, the absolute values of the derivatives are applied. These sensitivities are then multiplied by an assumed worst case input variation,

$$\Delta \dot{m}_{\text{nox,series}}(\mathbf{u}) = \left| \frac{\partial \dot{m}_{\text{nox}}(\mathbf{u})}{\partial \mathbf{u}} \right| \sigma_{\text{sensor}} \quad (4.18a)$$

$$\text{and } \Delta \dot{m}_{\text{soot,series}}(\mathbf{u}) = \left| \frac{\partial \dot{m}_{\text{soot}}(\mathbf{u})}{\partial \mathbf{u}} \right| \sigma_{\text{sensor}}. \quad (4.18b)$$

Variations can be due to sensor variances or construction tolerances. The input variations  $\sigma_{\text{sensor}}$  may therefore be determined by data sheets or empirical values and are here given as standard deviations for the inputs  $m_{\text{air}}$ ,  $p_{2i}$ ,  $\varphi_{Q50}$  and  $T_{2i}$  by

$$\sigma_{\text{sensor}} = [4, 0.02, 1, 2] \quad (4.19)$$

$$\text{in } \left[ \frac{\text{mg}}{\text{cyc}}, \text{bar}, ^\circ\text{CA}, ^\circ\text{C} \right].$$

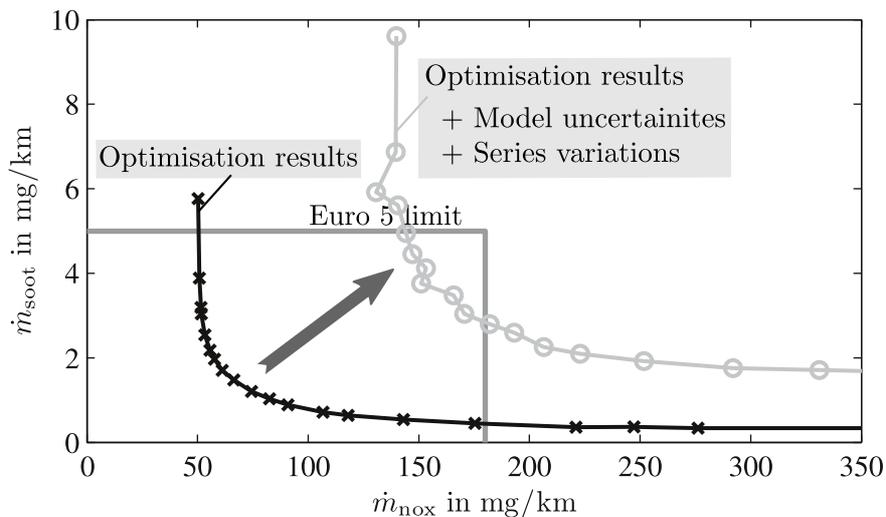
Note that due to the absolute value in eq. (4.18), positive and negative differences are regarded for the sensor variations in eq. (4.19). Hence the variations for the air mass flow rate covers  $\approx 2\%$  of the relevant measurement region. The variations for the intake pressure covers  $\approx 5\%$ , for the crank angle of main injection  $\approx 10\%$  and the intake temperature  $\approx 10\%$ . Since for the soot model the gas composition is applied as model input instead of the air mass flow rate, the variation in  $m_{\text{air}}$  is converted into a variation in  $x_{\text{eng,out}}$  for the soot model.

Applying eq. (4.18), the series variations can be determined for a regarded optimisation value  $\mathbf{u}$  in each operation point. The influences of the series variations for the operation point  $n_{\text{eng}} = 2500$  rpm

and  $u_{inj} = 30 \text{ mm}^3/\text{cyc}$  are indicated by the dashed grey line in Fig. 4.3. It is mainly a shift towards higher emissions and needs therefore be considered for the global optimisation presented in the following. These informations about sensitivities can also be applied to formulate requirements for sensor accuracies and construction tolerances.

### 4.3 Global optimisation

For a global optimisation, all  $M$  operating points are regarded simultaneously. The goal is to optimise the bsfc over the test cycle, which is equivalent to maximise the engine energy over the test cycle. The engine energy is determined by a weighted sum of the local engine torques, see eq. (4.14). Since for a global optimisation all local model inputs are optimised simultaneously, a high dimensional problem arises. The initial values are therefore taken from the local optimisation results. Since several weights  $k_{nox}$  and  $k_{soot}$  are applied for the local optimisation, there is a variety of initial values to select from. To choose appropriate initial values, the accumulated emissions for the NEDC are regarded with and without model uncertainties and series variations in Fig. 4.4. These are especially not the results of a global optimisation, but the accumulated emissions resulting from various local optimisations with varying weights.



**Figure 4.4:** Accumulated emissions for the NEDC without model uncertainties and without series variations (black) and with the worst case assumptions for model uncertainties and series variations (grey). This global Pareto front is calculated by the weighted sum of the local Pareto fronts. The local Pareto fronts are derived from the local optimisations with various weights.

The accumulated emissions are calculated by the weighted sum of the local emissions, see eq. (4.12). The black curve in Fig. 4.4 shows the accumulated emissions for the local optimisation results for various weights,  $k_{nox}$  and  $k_{soot}$ . Model uncertainties and series variations affect these optimisation results and the transferability to other engines. These local uncertainties sum up to a global uncertainty and deteriorate the accumulated raw emissions over a test cycle. The worst case assumptions

of these uncertainties are therefore added to the presented accumulated emissions, as is presented by the grey curve in Fig. 4.4. If the uncertainties are neglected, the optimised test engine fulfils the boundary conditions of the Euro 5 emission standard for several weights  $k_{\text{nox}}$  and  $k_{\text{soot}}$  (black crosses inside the Euro 5 limit). If the uncertainties are regarded, the Pareto front is shifted towards higher emissions and only few weights fulfil the Euro 5 emission standard (grey circles inside the Euro 5 limit). Initial values for the global optimisation are therefore selected such that the limits are fulfilled with and without consideration of uncertainties.

Fig. 4.4 shows that several optimisation results which are in the range without the consideration of uncertainties violate the emission standards if uncertainties are considered. It is therefore crucial to regard these uncertainties at an early state of optimisation, which is given by the robust optimisation presented in the following.

### 4.3.1 Robust optimisation

The goal of a robust optimisation is to provide feasible results, even if several variations occur. The regarded variations are the model uncertainties and the series variations, as presented in Sect. 4.2.2. Each local model has an individual model uncertainty and an individual series variation. These local variations sum up to a global variation with regard to the local weights  $k_j$ . The variations further depend on the the regarded optimal values  $\mathbf{u}_{\text{opt}}$ .

Besides the model uncertainties and series variations, there are additional emissions due to engine dynamics. The additional emissions depend on the applied control structure, which is an air mass control. This air mass control adjusts the intake pressure and the air mass flow rate by the actuators of the turbocharger  $s_t$  and the egr system  $s_{\text{egr}}$ . Furthermore, the intake temperature is adjusted by  $s_{\text{lp-egr}}$ . These and alternative control structures are further discussed in Chap. 5. To determine the additional dynamic emissions, the NEDC is simulated applying the air mass control for the air path states with the desired values taken from the local optimisation. The accumulated emissions are compared to the stationary simulated emissions and the differences between those give the additional dynamic emissions,

$$\Delta m_{\text{nox,dyn}} = 11 \text{ mg/km} \quad (4.20a)$$

$$\Delta m_{\text{soot,dyn}} = 0 \text{ mg/km}. \quad (4.20b)$$

The applied air mass control results in additional  $\text{NO}_x$  emissions and avoids additional soot emissions, which is why additional emissions are only considered for  $\text{NO}_x$ . In contrast, if an egr-rate control is applied, the additional emissions reverse, since an egr-rate control results in additional soot emissions, see Sect. 5.3. Then it would be necessary to add a dynamical delta as in eq. (4.20) to the soot emissions and not to the  $\text{NO}_x$  emissions.

To consider all these variations in the optimisation process, the variations are incorporated in the boundary conditions of eq. (4.12) by

$$\sum_{j=1}^M \frac{k_j (\dot{m}_{\text{nox},j}(\mathbf{u}_{\text{opt}}) + \Delta \dot{m}_{\text{nox},\text{mod},j}(\mathbf{u}_{\text{opt}}) + \Delta \dot{m}_{\text{nox},\text{series},j})}{s_{\text{cycle}}} + \Delta m_{\text{nox},\text{dyn}} \leq \text{NO}_{\text{x},\text{limit}} \quad (4.21\text{a})$$

$$\sum_{j=1}^M \frac{k_j (\dot{m}_{\text{soot},j}(\mathbf{u}_{\text{opt}}) + \Delta \dot{m}_{\text{soot},\text{mod},j}(\mathbf{u}_{\text{opt}}) + \Delta \dot{m}_{\text{soot},\text{series},j})}{s_{\text{cycle}}} + \Delta m_{\text{soot},\text{dyn}} \leq \text{soot}_{\text{limit}} \quad (4.21\text{b})$$

Applying these additional terms in the emission constraint, the derived optimisation result keeps a safety distance to the emission limits. This distance arises from the model uncertainties, the worst case assumptions for series variations and the additional dynamic emissions. Optimisation results are shown in Sect. 4.4.

### 4.3.2 Smoothing of engine maps

Once the set points for the engine control are derived from the global optimisation, the drivability of these set points needs to be ensured. Because of the applied constraint in eq. (4.13), the derived set points are reachable in practice. Since the optimisation is performed for all  $M$  operating points, the optimised set points can be stored as look-up tables over these operation points. The elements of the optimised vector  $\mathbf{u}_{\text{opt}}$  are therefore stored as the desired values for air mass per cycle, intake pressure, crank angle of 50% mass fraction burnt and intake temperature,

$$m_{\text{air,des}} = m_{\text{air}}(n_{\text{eng}}, u_{\text{inj}}) \quad (4.22\text{a})$$

$$p_{2i,\text{des}} = p_{2i}(n_{\text{eng}}, u_{\text{inj}}) \quad (4.22\text{b})$$

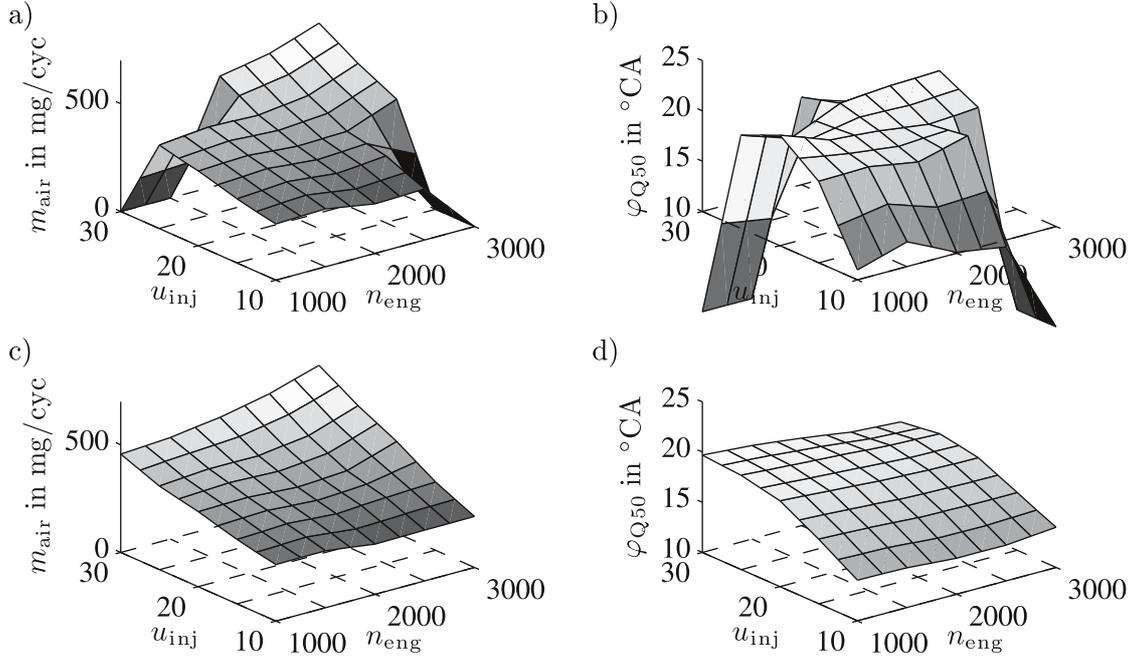
$$\varphi_{\text{Q50,des}} = \varphi_{\text{Q50}}(n_{\text{eng}}, u_{\text{inj}}) \quad (4.22\text{c})$$

$$T_{2i,\text{des}} = T_{2i}(n_{\text{eng}}, u_{\text{inj}}). \quad (4.22\text{d})$$

The look-up tables of  $m_{\text{air,des}}$  and  $\varphi_{\text{Q50,des}}$  are depicted in the upper two plots in Fig. 4.5. These are the model based optimisation results, which directly result from the global optimisation. Since some operation points in the corners are not regarded for optimisation, these are set to zero. Furthermore, for  $\varphi_{\text{Q50,des}}$  high variations occur also over the inner operation points. These variations cause unnecessary actuating energy and wear of materials. To reduce these effects and ensure a smooth engine running, the look-up tables are smoothed.

The look-up tables are smoothed with a regularisation for look-up tables as presented in Sect. 2.1.2. This regularisation penalises the curvature of a look-up table with regard to a regularisation parameter  $\lambda$ . The smoothness of the look-up tables can therefore be adjusted by a variation of this parameter. The regularisation for the look-up tables in Fig. 4.5 is determined by trial and error of several parameters and selection by visual decision.

Instead of smoothing the look-up tables after optimisation, a smoothing can also be integrated in the optimisation process. Therefore, either an implicit or an explicit regularisation can be applied.



**Figure 4.5:** Set points for the air mass per cycle  $m_{\text{air}}$  and the crank angle of 50% mass fraction burnt  $\varphi_{Q50}$  derived from the global optimisation. The plots a) and b) show the results without and the plots c) and d) with regularisation.  $u_{\text{inj}}$  is given in  $\text{mm}^3/\text{cyc}$  and  $n_{\text{eng}}$  in rpm.

For an implicit smoothing, a look-up table is parametrised by a functional expression with regard to the parameters  $\theta$ , e. g.

$$m_{\text{air}}(n_{\text{eng}}, u_{\text{inj}}) = f_{\theta}(n_{\text{eng}}, u_{\text{inj}}). \quad (4.23)$$

It is generally assumed that less parameters are applied for  $f_{\theta}$  than for the look-up table. Then the degrees of freedom are reduced and a regularisation is introduced. This also reduces the dimensionality of the optimisation problem, since the parameters  $\theta$  are optimised rather than the input vector  $\mathbf{u}$ . Problematic is however that the constraints need then also be reformulated with regard to this parametrisation. Furthermore, the interpretability of the parameters is lost and the analytic determination of derivatives is highly complicated. Additionally to this, there is no regularisation parameter to adjust the degree of smoothing.

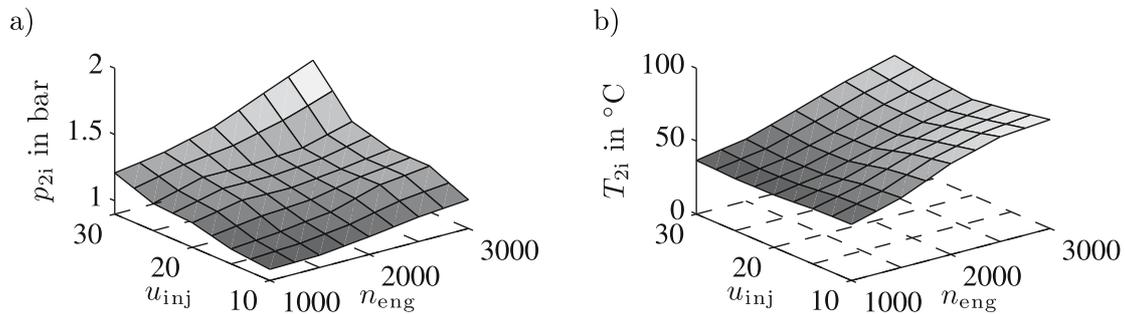
An explicit smoothing adds a penalty term to the objective function in eq. (4.11) such that

$$J(\mathbf{u}) = \text{bsfc}(\mathbf{u}) + \lambda \Gamma \mathbf{u}, \quad (4.24)$$

see Sect. 2.1.2. The choice of  $\lambda$  needs then be adapted to the optimisation problem by trial and error, for which several optimisation runs might be necessary. Because of these several runs and the high dimensional input of the global optimisation, the explicit smoothing highly increases the computing time. It would therefore be necessary to accelerate the optimisation algorithm before integrating such a regularisation to the optimisation problem. Therefore, the explicit smoothing is applied after the global optimisation to the optimised maps, see Fig 4.5 c) and d).

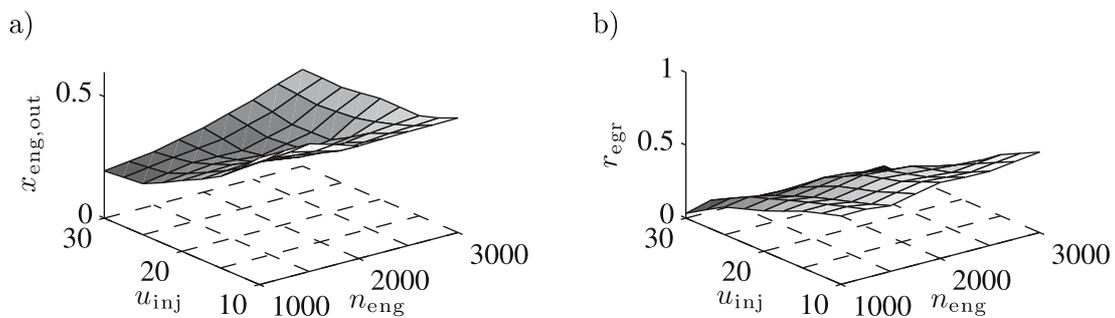
The smoothed results for the other inputs,  $p_{2i}$  and  $T_{2i}$  are shown in Fig. 4.6. The set points for the intake pressure increases with the injection quantity and the engine speed. For these operation

points the exhaust gas possesses a higher enthalpy, why the turbocharger can produce higher intake pressures. The increase in intake pressure also enables higher air mass flow rates as can be seen especially for higher injection quantities in Fig. 4.5 c).



**Figure 4.6:** Set points for the intake pressure  $p_{2i}$  and the intake temperature  $T_{2i}$  derived from the global optimisation.  $u_{inj}$  is given in  $\text{mm}^3/\text{cyc}$  and  $n_{eng}$  in rpm.

Given these set points it is also possible to determine further air path states which are redundant for stationary conditions. In Fig. 4.7 the determined air path states for  $x_{eng,out}$  and  $r_{egr}$  are shown. It can be seen that  $x_{eng,out}$  is relatively high for low injection quantities and decreases towards higher injection quantities. For high engine speeds it is almost constant. The egr rate is approximately constant over the engine speed and decreases linearly from low injection quantities to high injection quantities. For high injection quantities the egr rate is zero. The presented set points for  $x_{eng,out}$  and  $r_{egr}$  are alternative set points to  $m_{air}$  for a closed loop control of the air path.



**Figure 4.7:** Set points for the gas composition after combustion  $x_{eng,out}$  and the egr rate  $r_{egr}$ . For stationary conditions, these set points are redundant to the air path states presented in Fig. 4.5 and Fig. 4.6. They are therefore determined from these results.  $u_{inj}$  is given in  $\text{mm}^3/\text{cyc}$  and  $n_{eng}$  in rpm.

## 4.4 Discussion of optimisation results

The optimisation results of the local and the global optimisation are compared and discussed in the following. At first, differences between the local and the global optimisation results are regarded.

Then, variations of a series production are simulated and finally the stationary results are compared to a dynamic simulation of the NEDC.

Comparisons are given in bsfc and accumulated emissions over the NEDC. The accumulated emissions are calculated by the summed up emissions over the NEDC divided by the covered distance  $s_{\text{cycle}} = 11.014$  km. The bsfc is given in g/kWh and is determined by dividing the summed up injection quantity by the produced energy on the crankshaft, see eq. (4.11). The bsfc can alternatively be calculated with regard to the produced inner energy,  $\text{bsfc}_{\text{inner}}$ . The inner bsfc is determined by applying the inner torque without friction losses in eq. (4.11) instead of the torque on the crankshaft,  $M_{\text{eng,inner}} = M_{\text{eng}} - M_{\text{eng,friction}}$ . In contrast to the torque on the crankshaft, the inner torque can not become negative, why it is better suited for some comparisons. In the following, both bsfc's are stated for the optimisation results.

#### 4.4.1 Comparison of the local and global optimisation results

The results of the local optimisation are applied as initial values for the global optimisation. There are several local optimisation results to select from, which are derived from various weights  $k_{\text{nox}}$  and  $k_{\text{soot}}$ . The initial values are taken from the optimisation results with the weights derived from  $\tilde{k}_{\text{nox}} = 0.6$ . The accumulated emissions with these weights are inside the Euro 5 limit with and without the consideration of model uncertainties and series variations, see Fig. 4.4. The bsfc and the accumulated emissions for these results of the local optimisation, for the results of the global optimisation and the results of the robust global optimisation are stated in Tab. 4.1.

**Table 4.1:** Optimisation results given for the local optimisation, the global optimisation and the robust global optimisation. Accumulated emissions for NO<sub>x</sub> and soot and the bsfc are shown, calculated by the weighted sum as in eq. (4.12) and eq. (4.11) with weights derived from the NEDC. The inner bsfc,  $\text{bsfc}_{\text{inner}}$ , neglects the towing torque. Engine efficiencies are stated in parentheses.

	$\sum \dot{m}_{\text{nox}}$	$\sum \dot{m}_{\text{soot}}$	bsfc ( $\eta$ )	$\text{bsfc}_{\text{inner}} (\eta_{\text{inner}})$
local optimisation	61.2 $\frac{\text{mg}}{\text{km}}$	1.70 $\frac{\text{mg}}{\text{km}}$	314.7 $\frac{\text{g}}{\text{kWh}}$ (26.9 %)	179.6 $\frac{\text{g}}{\text{kWh}}$ (47.2 %)
global optimisation	180.0 $\frac{\text{mg}}{\text{km}}$	3.75 $\frac{\text{mg}}{\text{km}}$	291.7 $\frac{\text{g}}{\text{kWh}}$ (29.1 %)	171.8 $\frac{\text{g}}{\text{kWh}}$ (49.3 %)
robust global optimisation	91.2 $\frac{\text{mg}}{\text{km}}$	2.17 $\frac{\text{mg}}{\text{km}}$	304.3 $\frac{\text{g}}{\text{kWh}}$ (27.9 %)	176.1 $\frac{\text{g}}{\text{kWh}}$ (48.1 %)

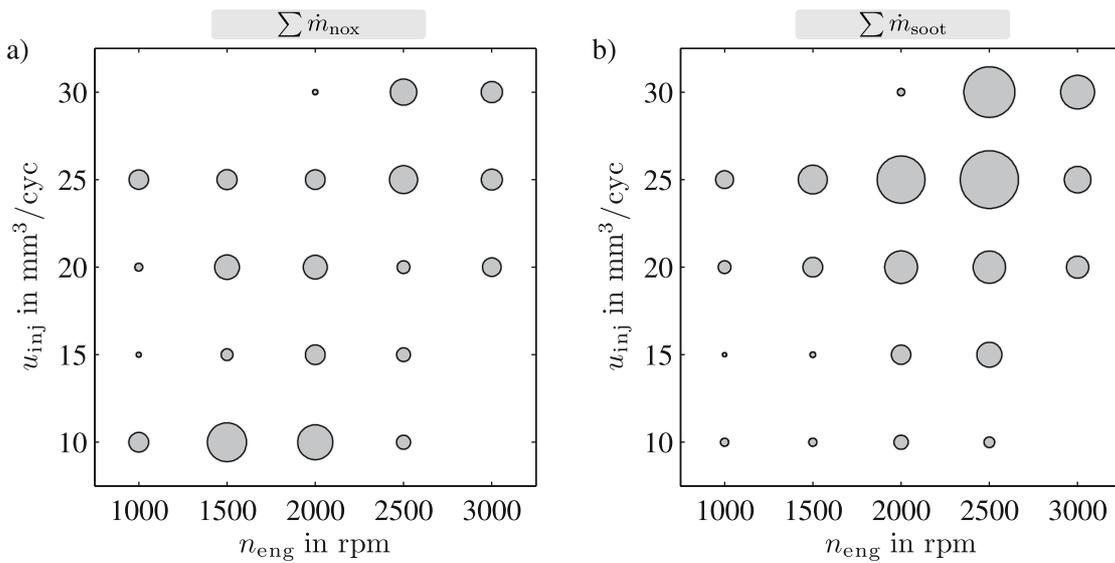
Since the engine torque is not regarded in the local objective function, see eq. (4.1), the local optimisation provides the worst results with regard to the bsfc. The accumulated emissions  $\sum \dot{m}_{\text{nox}}$  and  $\sum \dot{m}_{\text{soot}}$  on the other hand are well inside the Euro 5 limit. In contrast, the global optimisation minimises the bsfc but the accumulated NO<sub>x</sub> emissions are on the limit value. For the robust global optimisation, a safety distance is kept to the NO<sub>x</sub> and soot limits, such that a variation does not directly cause a violation of these limits. The robustness is on the cost of a higher fuel consumption, as is visible by the bsfc results. The additional fuel consumption for a robust optimisation is

about  $\approx 2.5\%$ , compared to the global optimisation. The stated efficiencies in parentheses are with regard to the theoretical energy content of Diesel fuel, which is defined by the lower heating value  $\text{LHV}_{\text{Diesel}} = 0.0118 \text{ kWh/g}$  [177]. The efficiencies are determined by

$$\eta = \frac{1}{0.0118 \frac{\text{kWh}}{\text{g}} \text{bsfc}} \quad (4.25a)$$

$$\text{and } \eta_{\text{inner}} = \frac{1}{0.0118 \frac{\text{kWh}}{\text{g}} \text{bsfc}_{\text{inner}}}. \quad (4.25b)$$

To analyse where the emissions come from, the produced emissions are regarded separately for each operation point. The accumulated emissions for an operation point are indicated by the sizes of the circles in Fig. 4.8. The left plot shows the produced  $\text{NO}_x$  emissions and the right plot the soot emissions. Despite the small weights in the engine operation points with high speed and high injection quantity, see Fig. 4.1, the accumulated emissions in these operation points are similar to the accumulated emissions in the operation points with low engine speeds and low injection quantities. It is therefore especially not feasible to limit the average emission mass flow rates for the local optimisation criterion, as suggested in eq. (4.10).

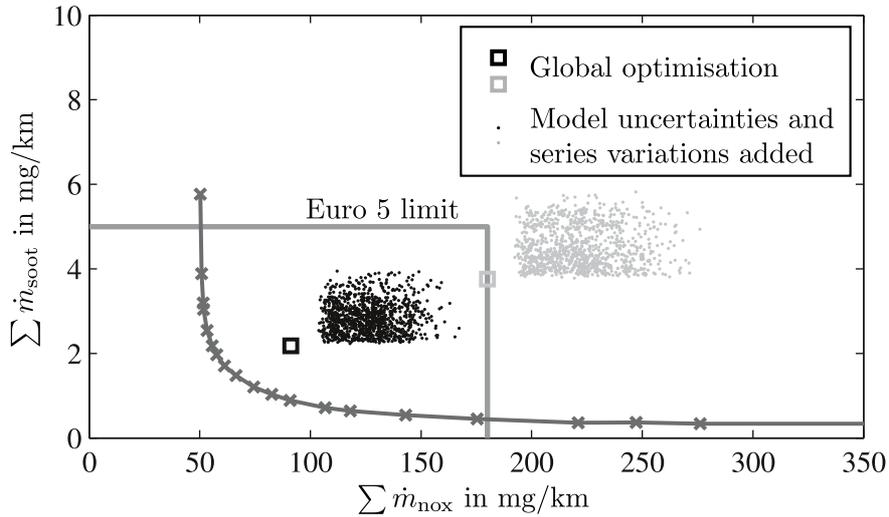


**Figure 4.8:** Distribution of emissions over the operation points for the results of the robust global optimisation with robustness constraints. Sizes of the circles indicate a) the accumulated  $\text{NO}_x$  emissions  $\sum \dot{m}_{\text{nox}}$  and b) the accumulated soot emissions  $\sum \dot{m}_{\text{soot}}$  over the NEDC for the various operation points.

#### 4.4.2 Simulation of an engine series production

For an investigation of an engine series production, the influences of variations due to model uncertainties, series tolerances and dynamic emissions are regarded. Therefore, the results of the global

optimisation are investigated with and without the consideration of these variations. The grey square in Fig. 4.9 shows the emissions derived from a global optimisation with constraints as in eq. (4.12), and the black square shows the results from a robust global optimisation with constraints as in eq. (4.21). These results correspond to the results stated in Tab 4.1. Both fulfil the Euro 5 emission standard, but the global optimisation results are on the  $\text{NO}_x$  limit, whereas the robust global optimisation results have a safety distance to the limit values.



**Figure 4.9:** Accumulated emissions for the NEDC with set points being derived from the global optimisation with robustness constraints (black) and without robustness constraints (grey), see eq. (4.21) and eq. (4.12) respectively. 1000 Monte-Carlo simulations of uncertainties are shown by the dots. The variations are assumed to be normally distributed and a fixed offset is added to the  $\text{NO}_x$  emissions accounting for the additional dynamic emissions, see eq. (4.12).

To simulate the accumulated emissions for an engine series production, the model uncertainties, series variations and dynamic emissions are added to the optimised values. The dynamic emissions, see eq. (4.20), are added as a constant offset to the optimised values. The model uncertainties and the series variations are simulated by a Gaussian distribution. These variations are limited to positive values such that variations can only deteriorate and not improve the optimisation results. 1000 Monte-Carlo simulations are performed to simulate a series production, see Fig. 4.9. For the results of the global optimisation, the Euro 5 emission standard is violated for every realisation, see the grey dots in Fig. 4.9. This is since the fuel consumption is optimized on cost of the emissions. In contrast, for the robust global optimisation all of the realisations are in the range of the limits, see the black dots in Fig. 4.9.

The influences of the individual variations can further be analysed, by investigating the individual contribution of each variation. The worst case contributions of the model uncertainties, series variations and the additional dynamic emissions are listed in Tab. 4.2. The series variations are determined by a multiplication of the model sensitivities and the variation vector  $\sigma_{\text{sensor}}$ , see eq. (4.18). These variations are therefore further divided into the individual contributions of the inputs.

**Table 4.2:** Contributions to the robust optimisation constraint of eq. (4.21) for the results of the robust global optimisation. Worst case contributions are shown with regard to the NEDC in mg/km.

	$\Delta\dot{m}_{\text{nox}}$	$\Delta\dot{m}_{\text{soot}}$
model uncertainties	22.3 $\frac{\text{mg}}{\text{km}}$	0.49 $\frac{\text{mg}}{\text{km}}$
series variations $m_{\text{air}}$	13.5 $\frac{\text{mg}}{\text{km}}$	0.60 $\frac{\text{mg}}{\text{km}}$
$p_{2i}$	19.2 $\frac{\text{mg}}{\text{km}}$	0.32 $\frac{\text{mg}}{\text{km}}$
$\varphi_{Q50}$	19.0 $\frac{\text{mg}}{\text{km}}$	0.63 $\frac{\text{mg}}{\text{km}}$
$T_{2i}$	3.9 $\frac{\text{mg}}{\text{km}}$	0.10 $\frac{\text{mg}}{\text{km}}$
dynamic emissions	11.0 $\frac{\text{mg}}{\text{km}}$	0.00 $\frac{\text{mg}}{\text{km}}$
Total	88.8 $\frac{\text{mg}}{\text{km}}$	2.14 $\frac{\text{mg}}{\text{km}}$

The contributions of the model uncertainties are in the range of the largest contributions of the series variations. The biggest contributions among the series variations are given by  $p_{2i}$  for the  $\text{NO}_x$  emissions and by  $\varphi_{Q50}$  for the soot emissions. An analysis of the individual contributions of the series variations allows conclusions about the required accuracies for the individual inputs. Thus, the transferability of the optimisation results from a test engine to a series production can be evaluated. Furthermore, the results in Tab. 4.2 give an insight about the relative proportions of the uncertainties and variations to the accumulated emissions. The total sum of all variations and uncertainties illustrates the difficulty of transferring the optimisation results from a test engine to a series of engines. For the regarded case, the total sum for the worst case assumptions is in the range of the accumulated emissions derived from the robust global optimisation, compare Tab. 4.1 and Tab. 4.2. It further shows that the contributions of the dynamic emissions are relatively small compared to the other variations.

Note that the addition of the total  $\text{NO}_x$  value in Tab. 4.2 and the accumulated  $\text{NO}_x$  emissions derived from the robust global optimisation in Tab. 4.1 is again on the  $\text{NO}_x$  limit. Hence, if the distance to the emission limits shall be increased or decreased, this can be adjusted via the robust optimisation constraints formulated in eq. (4.21).

### 4.4.3 Dynamic simulation of the NEDC

For the presented stationary optimisation, the bsfc and the accumulated emissions are calculated by a weighted sum over 21 operation points, see eq. (4.11) and eq. (4.12). The weights  $k_j$  for the operation points are derived from the frequency of occurrence of the operation points in the NEDC. Thus, the accumulated emissions and the bsfc are calculated in a reasonable computing time. In the following, this *local weighted summation* is compared to a *quasistationary simulation* and to a *dynamic simulation* of the NEDC.

For the quasistationary simulation, the entire NEDC is simulated with the assumption that all dynamic effects are decayed at all time. This is similar to the local weighted summation, since the local weighted summation is the sum of the occurrence of the operation points assuming stationary conditions. Differences are given because of the simplification to 21 operation points, instead of regarding the entire operation range. The quasistationary simulation requires a longer calculation time, since the  $t_{\text{cycle}} = 1180$  s of the NEDC are discretised in  $\Delta t = 0.5$  s time steps. Hence, the emissions and the bsfc are calculated for 2160 stationary operation points.

For the dynamic simulation of the NEDC, a dynamic air path model is connected to the combustion model, as depicted in Fig. 3.1. This overall engine model is computed with a discretisation of  $\Delta t = 0.005$  s such that also fast dynamic effects of the air path system are describable. For the desired optimised air path set points  $m_{\text{air}}$ ,  $p_{2i}$  and  $T_{2i}$ , closed loop controls are utilised. The crank angle of 50% mass fraction burnt  $\varphi_{Q50}$  is controlled by an open loop control of the crank angle of main injection  $\varphi_{\text{mi}}$ .

The accumulated emissions and the bsfc are summarised for the three approaches in Tab. 4.3. The emissions are almost identical for the local weighted summation and the quasistationary simulation. Hence, the simplification to a weighted summation is feasible. The differences in bsfc are due to the operation points with low load for which the engine is dragged. These operation points are not regarded for the local weighted summation, why the negative torques in these operation points only influence the quasistationary and the dynamic simulation. Regarding  $\text{bsfc}_{\text{inner}}$ , the differences are significantly smaller, since no negative torques are produced here.

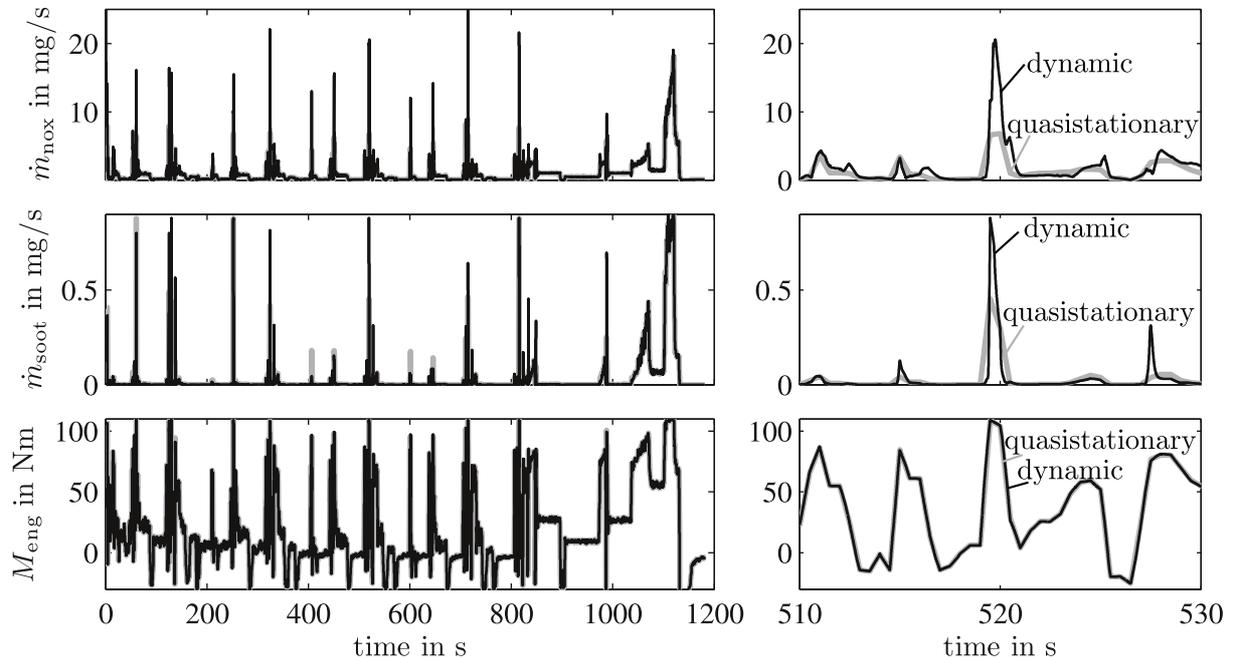
**Table 4.3:** Accumulated emissions and bsfc calculated by a weighted sum of local operation points, by a quasistationary simulation of the NEDC and by a dynamic simulation of the NEDC. Set points are taken from the robust global optimisation with a smoothing of look-up tables.

	$\sum \dot{m}_{\text{nox}}$	$\sum \dot{m}_{\text{soot}}$	bsfc ( $\eta$ )	bsfc <sub>inner</sub> ( $\eta_{\text{inner}}$ )
local weighted summation	97.8 $\frac{\text{mg}}{\text{km}}$	2.96 $\frac{\text{mg}}{\text{km}}$	306.5 $\frac{\text{g}}{\text{kWh}}$ (27.7 %)	176.9 $\frac{\text{g}}{\text{kWh}}$ (47.9 %)
quasistationary simulation	99.2 $\frac{\text{mg}}{\text{km}}$	3.14 $\frac{\text{mg}}{\text{km}}$	432.1 $\frac{\text{g}}{\text{kWh}}$ (19.6 %)	183.7 $\frac{\text{g}}{\text{kWh}}$ (46.1 %)
dynamic simulation	109.4 $\frac{\text{mg}}{\text{km}}$	2.99 $\frac{\text{mg}}{\text{km}}$	432.7 $\frac{\text{g}}{\text{kWh}}$ (19.6 %)	184.5 $\frac{\text{g}}{\text{kWh}}$ (45.9 %)

Regarding the dynamic simulation, the bsfc differs only slightly from the quasistationary simulation. This is since the torque is mainly influenced by the injection quantity, which is not affected by any dynamics. There are however differences observable for the dynamic simulated emissions, which are due to the dynamics of the air path. Again, a closed loop control is applied for the air path states. The applied air mass control favours the formation of  $\text{NO}_x$  and avoids soot emissions, which can be seen by the increased  $\text{NO}_x$  emissions and the almost unaffected soot emissions for the dynamic simulation.

The additional dynamic emissions are in the range of the assumed dynamic variation for the robust optimisation, see eq. (4.20). The additional emissions are relative small, since changes in the operation point are implemented as ramps instead of steps. Thus, a smooth transitions from one

operation point to another is enabled such that the air path states follow their desired set points with small deviations. A comparison between the quasistationary simulation and the dynamic simulation is given in Fig. 4.10 for the entire NEDC on the left and a zoomed section on the right.



**Figure 4.10:** Quasistationary and dynamic simulation of the emissions and the torque for the NEDC with set points derived from the robust global optimisation. For the quasistationary simulation it is assumed that all dynamic effects are decayed such that these are simulated with the combustion model alone. The dynamic simulation is given by the connection of the combustion model with the dynamic air path model as depicted in Fig. 3.1.

The torque of the dynamic simulation is almost identical to the quasistationary simulation, since the torque is only little affected by the dynamics of the air path. The dynamic emissions are however more affected by these dynamics. The dynamic emissions can be increased or decreased compared to their stationary emissions, often depending on the direction of the dynamic. For example, the dynamic  $\text{NO}_x$  emissions are lower than their quasistationary value for a engine speed ramp up, whereas the dynamic emissions are higher than their quasistationary emissions for a ramp down, compare to Fig. 3.13. Hence, the increased dynamic emissions are partially compensated by the decreased dynamic emissions. This lowers the difference between the accumulated quasistationary and the accumulated dynamic emissions.

## 4.5 Summary

The model-based optimisation of the stationary engine operation is presented. The optimisation is divided into a local and global optimisation. The local optimisation regards each engine operation

point separately, whereas the global optimisation optimises all engine operation points simultaneously with regard to a given test cycle.

The objective function for the local optimisation is a weighted sum of the  $\text{NO}_x$  and soot emissions. The weights of the objective function are varied such that a Pareto front results, showing the  $\text{NO}_x$ -soot trade-off. The quality of the obtained optimisation results depends on the applied numerical optimisation algorithm. Local optimisation algorithms depend on the quality of the initial value and can get stuck in a local optimum. Global optimisation algorithms are more reliable in finding the global optimum but require a higher computing time.

The objective function for the global optimisation is directly derived from the legislative boundaries and the customer demands. The goal is to minimise the brake specific fuel consumption over the regarded test cycle under the constraints of the legislative emission limits. The minimisation of the brake specific fuel consumption is equivalent to a maximisation of a weighted sum of the local engine torques. The combustion models as presented in Chap. 3 are applied for the optimisations.

Results obtained for a single test engine can not directly be transferred to a series of engines, since series variations and construction tolerances deteriorate the optimisation results. Furthermore, predicted emissions might be too optimistic due to model uncertainties. In addition to this, the stationary optimisation does not regard the dynamic emissions, which result from the dynamics of the air path system. To consider these variations, a robust global optimisation is introduced. The obtained results have a safety distance to the legislative limits such that the accumulated emissions are even for variations inside the demanded limits. This safety distance is at the expense of an increased fuel consumption by about  $\approx 2.5\%$ . A separation of the influences of the individual variations gives an insight about the individual contributions and allows to formulate requirements for sensor accuracies and construction tolerances.

Furthermore, the dynamic emissions over the NEDC are compared to the quasistationary emissions. To simulate the NEDC dynamically, the combustion model is connected to the dynamic air path model. With the given closed loop control of the air path states, the additional dynamic emissions are relatively small for the NEDC. This is because of the moderate dynamics of the NEDC and since dynamic emissions can be lower or higher compared to their quasistationary values, depending on the direction of the transient. However, the optimisation of the air path control allows further improvements as is considered in the next chapter.

---

## 5 Optimisation of the Dynamic Engine Operation

---

Due to the dynamics of the air path, the dynamic emissions differ from their stationary calibrated values. This is especially the case for acceleration events, why the air path actuator trajectories are optimised in the following for a typical acceleration event. The regarded objective function penalises the quadratic deviations of the dynamic emissions from their stationary calibrated values. Due to the relatively high computing time, such an optimisation can only be calculated offline. For an optimisation of the emissions during the engine operation, various closed loop controls are discussed. A coupled actuator control is introduced, which reduces the dynamics for a deceleration event. To further reduce the dynamic soot emissions, a model-based smoke limitation is presented. It limits the injection quantity during transients such that a predefined soot limit is not exceeded.

---

A great effort is taken during the engine development process to optimise the engine control with regard to fuel consumption and legislative emission limits. This is done in a first step for stationary engine operation and in a second step for dynamic engine operation. The optimisation of the dynamic engine operation becomes more and more important for future emission standards. If the optimised set points are adjusted by open loop controls, long settling times and additional emissions occur, see Fig. 3.13. The acceleration of these dynamics and the avoidance of additional emissions is the objective of a dynamic optimisation.

For a simulation of the dynamic engine operation, the stationary combustion model is connected to the dynamic air path model, as depicted in Fig. 3.1. The dynamic air path model covers the major dynamics and a connection to the combustion model simulates the overall engine behaviour from the engine actuators to the combustion outputs. It is again assumed that the fuel injection system possesses no dynamics, why it is not required for the overall engine simulation.

An objective function for a dynamic optimisation is defined in Sect. 5.1. It penalises the deviations of the dynamic emissions from their stationary calibrated values. Exemplary for an engine transient, an acceleration event is optimised in Sect. 5.2. The optimisation variables are the discretised trajectories of the air path actuators. Depending on the discretisation and the regarded time horizon, a high dimensional and computationally intensive optimisation problem arises. Hence, such an optimisation can only be performed offline and not be calculated online on the ECU during engine operation. It gives however a benchmark for closed loop controls and can be applied to design feedforward controls.

There are various approaches to determine feedforward and closed loop controls. The design of these is a complex task and the objective of several previous works and still in the interest of

current research projects. An overview of closed loop control is given in Sect. 5.3. Most of the control structures consider the air path alone, since the air path covers the major dynamics, others control the emissions directly. Furthermore, a coupled actuator control is introduced to reduce the  $\text{NO}_x$  emissions for a deceleration event.

The presented closed loop controls accelerate the dynamics of the air path, but the dynamics of the injection system are still considerably faster. A step in the acceleration pedal can therefore produce low air to fuel ratios, since the increase in injection can be realised instantaneously, whereas the increase in fresh air follows with a dynamic time lag. These low air to fuel ratios lead to the formation of soot, which is visible as smoke and poorly affects the accumulated emissions. These additional emissions can be avoided by limiting the injection quantity with regard to the current air path states, which is known as *smoke limitation*. Since this also reduces the engine torque, such a limitation should not be too stringent. In Sect. 5.4 a *model-based smoke limitation* is presented. It regards the high dimensional influences on the formation of soot by applying the soot model presented in Chap. 3 and requires no additional calibration effort.

## 5.1 Objective function

Stationary optimisation results are compromises of several objectives, such as fuel consumption, emissions, smooth engine running and others. Furthermore, the stationary optimisation results are often fine tuned on the engine test bed such that the derived settings are optimal with regard to a complex subjective optimisation criterion. It is therefore in general not possible to derive an objective function for the dynamic optimisation from the objective function of the stationary optimisation. Instead, the objective function for the dynamic optimisation can be formulated with regard to the stationary calibrated values. Since the emissions and the torque play a major role in the engine optimisation process, the emissions are regarded for this objective function.

The differences between the dynamic emissions,  $\dot{m}_{\text{nox}}(\mathbf{z}(t), \mathbf{s}(t))$  and  $\dot{m}_{\text{soot}}(\mathbf{z}(t), \mathbf{s}(t))$ , and the stationary calibrated values,  $\dot{m}_{\text{nox,stat}}(\mathbf{z}(t))$  and  $\dot{m}_{\text{soot,stat}}(\mathbf{z}(t))$ , are regarded.  $\mathbf{z}(t) = (n_{\text{eng}}(t), u_{\text{inj}}(t))$  is the engine operation point at time  $t$  and  $\mathbf{s}(t)$  denotes the positions of the actuators, see Fig. 3.1. Note that the stationary calibrated emissions depend only on the current engine operation point, since all dynamic effects are decayed for these values. To come up for the different magnitudes in  $\text{NO}_x$  and soot emissions, the differences are weighted by their stationary calibrated values,

$$J_{\text{dynamic}}(\mathbf{z}(t), \mathbf{s}(t)) = \int_0^{T_{\text{end}}} \left( \frac{\dot{m}_{\text{nox}}(\mathbf{z}(t), \mathbf{s}(t)) - \dot{m}_{\text{nox,stat}}(\mathbf{z}(t))}{\dot{m}_{\text{nox,stat}}(\mathbf{z}(t))} \right)^2 + \left( \frac{\dot{m}_{\text{soot}}(\mathbf{z}(t), \mathbf{s}(t)) - \dot{m}_{\text{soot,stat}}(\mathbf{z}(t))}{\dot{m}_{\text{soot,stat}}(\mathbf{z}(t))} \right)^2 dt. \quad (5.1)$$

Hence the minimisation of eq. (5.1) enforces a fast dynamic settling to the stationary calibrated emissions.

In the following, the produced  $\text{NO}_x$  and soot emissions are regarded for a typical acceleration event. For this acceleration event, the trajectories of the guide vanes of the turbocharger  $s_t$  and the egr valve  $s_{\text{egr}}$  are optimised,

$$\min_{\mathbf{s}(t)} J_{\text{dynamic}}(\mathbf{z}(t), \mathbf{s}(t)) \quad (5.2)$$

$$\text{with } \mathbf{s}(t)^T = [s_t(t), s_{\text{egr}}(t)]. \quad (5.3)$$

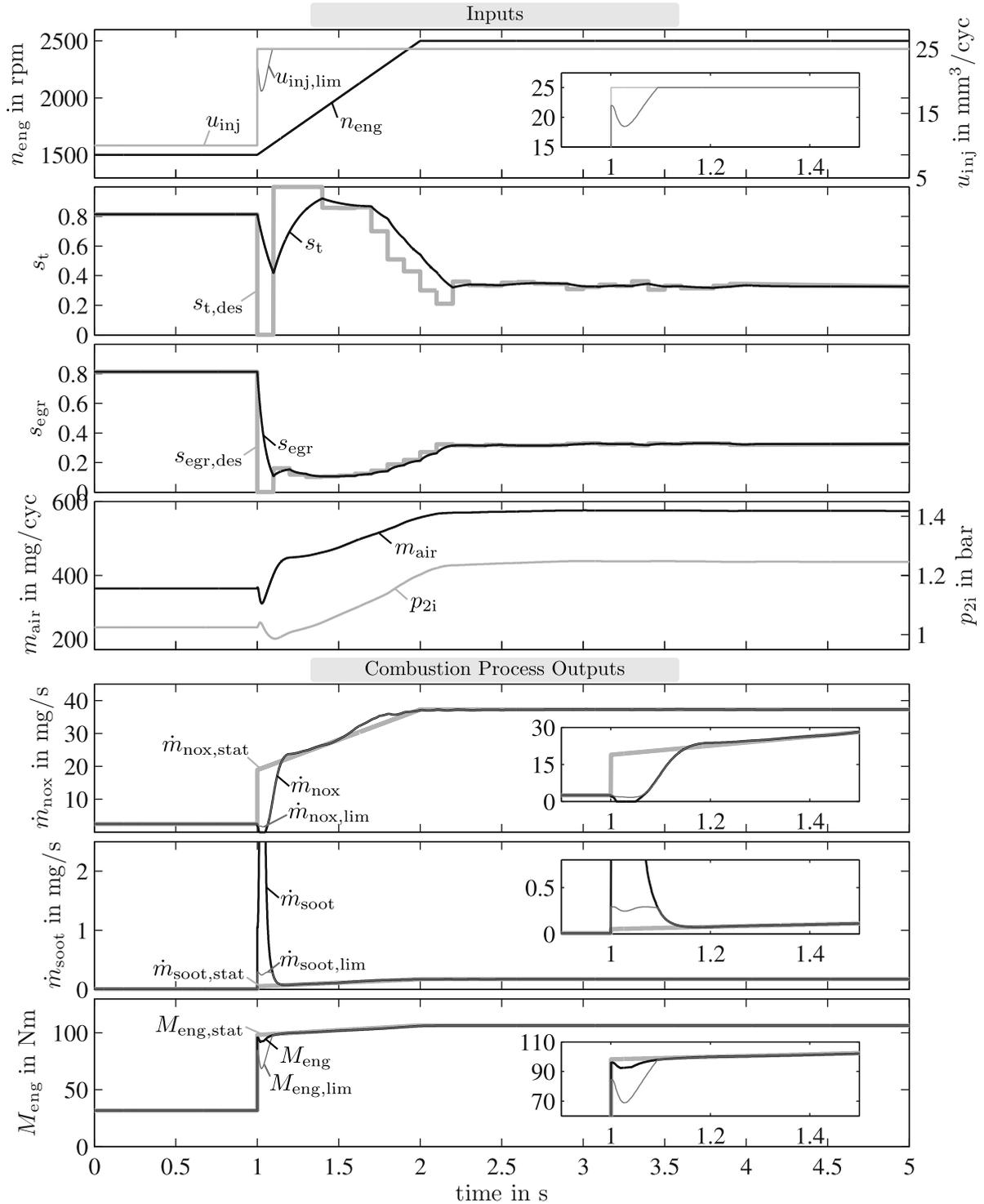
It is also possible to regard further engine actuators, such as the crank angle of main injection  $\varphi_{\text{mi}}$ , for this optimisation. Furthermore, additional terms can be regarded in the objective function, such as the engine torque. However, since the main dynamics result from the engine air path system and since these mainly influence the emissions  $\text{NO}_x$  and soot, the actuators as in eq. (5.3) are applied to optimise the objective function as in eq. (5.1).

## 5.2 Offline optimisation of an acceleration event

The acceleration event is defined as a step in injection quantity followed by a ramp in engine speed, see topmost plot in Fig. 5.1. From this predefined trajectory, the quasistationary emissions and the quasistationary torque result, depicted by the dashed lines in the lower three plots. The goal of the optimisation is that the dynamic emissions follow their quasistationary emissions, see eq. (5.1). The optimisation variables are the trajectories of the actuators  $s_t$  and  $s_{\text{egr}}$ .

The optimisation of the actuator trajectories requires a discretisation and an optimisation of each discretised time step. For a discretisation in  $N$  time steps and the two regarded actuators, the problem is of  $2N$ -dimensionality. Hence, already for a rough discretisation the dynamic optimisation is a high dimensional optimisation problem. The actuator trajectories are discretised in  $N = 30$  steps for the optimisation period of  $T_{\text{end}} = 3$  s. Such a discretisation is in the range of the actuator dynamics, why a finer discretisation improves the results only marginally. The discretisation can be regarded as a trade-off between the degrees of freedom for the actuator trajectories and the computing time required for optimisation. To achieve good initial values, the optimisation is at first performed with a discretisation in  $N = 10$  steps and narrowed thereafter.

The optimised values are shown by the dashed lines in the second and third plot in Fig. 5.1. Due to the actuator dynamics, the actuators follow their desired values with a dynamic lag, see the black lines in the second and third plot. The desired values are therefore overshoot to enforce a faster settling. It is interesting to see, that the guide vanes of the turbocharger are opened ( $s_t = 0$ ) for the first optimisation step, even though a higher intake pressure need to be attained for the quasistationary value. This opening of the guide vanes is advantage for the first few milliseconds, since it reduces the backpressure on the turbine such that less exhaust gas recirculates through the egr valve. After the first 100 ms, the egr valve is almost closed, such that the guide vanes of the turbocharger are closed again to build up the intake pressure. Again, the set point is overshoot to accelerate the dynamic settling. Note that the egr valve is not fully closed during the transient to control the  $\text{NO}_x$  emissions.



**Figure 5.1:** Optimisation of the air path actuator trajectories for an acceleration event given as a step in injection quantity  $u_{inj} = 10 \rightarrow 25 \text{ mm}^3/\text{cyc}$  followed by a ramp in engine speed  $n_{eng} = 1500 \rightarrow 2500 \text{ rpm}$  (topmost plot). Optimised actuator set points are shown by the dashed lines for  $s_t$  and  $s_{egr}$ . Actuator positions (black lines) follow with an actuator dynamic. The desired quasistationary combustion outputs (dashed grey lines) are shown in the lower plots together with the simulated dynamic combustion outputs (black lines). Combustion outputs with a smoke limitation are also shown (dark grey lines).

The stationary conditions are reached after 1.1 s, compared to the 3 s for the not optimised open loop control. The emissions for the not optimised open loop control are shown by the last transient in Fig. 3.13. A comparison between the optimised trajectories, the not optimised open loop control and a decentralised PID closed loop control of the air path states is given in Tab. 5.1. Average values over the acceleration event are shown for  $\text{NO}_x$ , soot and the engine torque. Values are given with regard to the quasistationary emissions, which are stated as 100 %. The applied closed loop control is again an air mass control, which controls the intake pressure and the air mass flow rate by a decentralised PID control structure.

**Table 5.1:** Averaged values for the emissions and the engine torque over an acceleration event ( $u_{\text{inj}} = 10 \rightarrow 25 \text{ mm}^3/\text{cyc}$  and  $n_{\text{eng}} = 1500 \rightarrow 2500 \text{ rpm}$ ). Results are shown for a quasistationary simulation, for a not optimised open loop control of the air path states, for a closed loop control of the air path states and for the optimised trajectories of the air path actuators as presented in Fig. 5.1.

	$\bar{m}_{\text{nox}}$	$\bar{m}_{\text{soot}}$	$\bar{M}_{\text{eng}}$
quasistationary	34.18 mg/s (100 %)	0.15 mg/s (100 %)	105.2 Nm (100 %)
open loop control	81.5 %	245.2 %	98.4 %
closed loop control	97.2 %	195.7 %	99.8 %
optimised trajectories	98.9 %	153.7 %	99.7 %

The  $\text{NO}_x$  emissions are in all cases less compared to the quasistationary emissions. This is because of the lack of fresh air for the first few milliseconds. This lack of fresh air also causes additional soot emissions such that a soot peak is observable, which poorly affects the averaged soot emissions. This soot peak is also observable for the optimised trajectories in Fig. 5.1. It is however significantly shorter than for the not optimised case, as shown in Fig. 3.13. The optimised trajectories reach stationary conditions the fastest, why the additional soot emissions are the smallest for this case. The not optimised open loop control produces the most soot emissions and the closed loop control is in between. The averaged torque is only little affected for all cases. This is because the torque mainly depends on the injection quantity.

The results of the optimised trajectories are the closest to the quasistationary emissions. There is however a significant difference for the soot emissions, which can no further be reduced with the air path actuators. A possibility to further reduce the soot emissions is by a smoke limitation, which limits the injection quantity with regard to the current air path states. Thus, the soot emissions but also the engine torque are reduced. The reduction in torque is relatively small, compared to the reduction in soot emissions. The emissions and the torque with a smoke limitation are indicated by the dark grey lines in the lower three plots in Fig. 5.1. The limited injection quantity can be seen in the topmost plot. The averaged values of the emissions and the torque over the acceleration event with a smoke limitation are listed in Tab. 5.2.

**Table 5.2:** Averaged values for the emissions and the engine torque over an acceleration event ( $u_{inj} = 10 \rightarrow 25 \text{ mm}^3/\text{cyc}$  and  $n_{eng} = 1500 \rightarrow 2500 \text{ rpm}$ ) with a smoke limitation. Quasistationary results are the same as the results without smoke limitation in Tab. 5.1.

	$\bar{m}_{nox}$	$\bar{m}_{soot}$	$\bar{M}_{eng}$
quasistationary	34.18 mg/s (100 %)	0.15 mg/s (100 %)	105.2 Nm (100 %)
open loop control	81.6 %	152.1 %	97.3 %
closed loop control	97.3 %	113.6 %	98.8 %
optimised trajectories	99.0 %	106.1 %	99.3 %

The smoke limitation reduces the soot emissions for all three cases significantly. Thus, the results of the optimised trajectories vary only little from the quasistationary simulation. The reduction in torque is in all cases relatively small, see also light grey line for  $M_{eng}$  in Fig. 5.1. Hence, the significant reduction in soot pays off the relatively small reduction in torque.

The results of the optimised trajectories can be regarded as a benchmark for the design of a closed loop control. Hence, there is some optimisation potential for the applied closed loop control, like the opening of the guide vanes of the turbocharger for the first few milliseconds of the acceleration event. This can not be achieved with the applied PID control structure, but might be incorporated into a feedforward control. The design of closed loop controls is an important part of the engine calibration process, but goes beyond the scope of this work. Therefore, an overview of various closed loop controls is given in the following.

### 5.3 Closed loop controls

The dynamic optimisation, as presented in the previous section, requires a large computing time and knowledge about the engine operation point trajectory. Since the trajectory is determined by the drivers request and the vehicle speed and since the computing capacity of the ECU is limited, such a dynamic optimisation can not be calculated online on the ECU. However, the optimisation provides a benchmark for the design of closed loop controls. Closed loop controls are applied during engine operation for the transition from one stationary operation point to another. There are various possibilities to design a control structure, which are shortly summarised in the following.

The regarded controls mainly concern the emissions  $\text{NO}_x$  and soot, which differ for transient operation from their stationary calibrated values, but also the engine torque, which is limited by the available air mass in the cylinder. The dynamic deviations are mainly due to the air path system, in particular due to the inertia of the turbocharger and the volume of the exhaust gas recirculation path. Since the turbocharger and the egr system have coupling effects on each other, a demanding control problem arises.

### Direct control of emissions

An intuitive approach is to control the emissions directly. In [8, 170] the  $\text{NO}_x$  concentration  $c_{\text{nox}}$  and the air to fuel ratio  $\lambda$ , representing the soot emissions, are controlled by the actuators  $s_{\text{egr}}$  and  $\varphi_{\text{mi}}$ . Instead of controlling  $\lambda$ , a virtual soot sensor is applied in [169] controlled by the swirl flaps  $s_{\text{sf}}$ . In [160] a coupled actuator structure is presented. One coupled actuator, combining  $s_{\text{egr}}$  and  $\varphi_{\text{mi}}$ , is employed to control  $c_{\text{nox}}$ , and the other, combining  $p_{\text{rail}}$  and  $s_{\text{t}}$ , to control soot measured as opacity.

The direct control of emissions has the drawback that the emissions need to be measured during engine operation. These measurements suffer from significant dynamics, see Chap. 3, and are expensive for a series production. Furthermore, state of the art soot sensors are maintenance intensive and not suited for a series application. Therefore, virtual sensors based on emission models or alternative control variables, such as the air to fuel ratio  $\lambda$  [8, 170], are applied.

### Air path controls

Alternatively to the direct control of emissions, the air path states can be controlled. Since the emissions depend on the air path states, this controls the emissions indirectly. Like for the inputs of the combustion model, discussed in Sect. 3.1.1, various air path states are suited as control variables. Furthermore, there are different air path configurations which demand different complex control structures.

A common air path configuration for passenger cars employs a turbocharger and an egr system. For such a configuration, two control variables are required. Common air path controls utilise the intake pressure  $p_{2i}$  as one of these control variables. The intake pressure can be measured dynamically fast with relative low priced series sensors and it is mainly affected by the turbocharger  $s_{\text{t}}$ . The intake pressure is often applied together with the air mass flow rate  $\dot{m}_{\text{air}}$  [143, 174] or with the egr rate  $r_{\text{egr}}$  [51, 141, 175]. The air mass flow rate is measured by the hot film air mass meter and the egr rate is calculated via an approximation of the exhaust gas flow rate  $\dot{m}_{\text{egr}}$ , which can also be applied as control variable [140]. Less common control structures utilise the exhaust gas pressure  $p_3$  [9, 175] or the gas composition in the intake  $x_{2i}$  and the exhaust system  $x_{\text{eng,out}}$  [5]. More advanced configurations employ a low-pressure egr system [45, 53, 108] and utilise additional control variables, such as the intake temperature.

For stationary operation, the various control variables lead to equivalent emissions. However, since a control of the air path states controls the emissions indirectly, the dynamic emissions vary for different control variables. For example, a control of  $p_{2i}$  and  $\dot{m}_{\text{air}}$  avoids dynamic soot emissions at the expense of additional  $\text{NO}_x$ , while the control of  $p_{2i}$  and  $r_{\text{egr}}$  avoids dynamic  $\text{NO}_x$  emissions at the expense of additional soot [196]. Comparisons with regard to various control variables are given in [52, 140, 175, 196].

The desired set points for the control variables are derived from a stationary optimisation. The set points are usually stored as two-dimensional look-up tables over the engine operation point. It is

also possible to correct the control variables for dynamic operation [107, 196] or to extend the look-up tables by a third dimension describing the degree of dynamic [115]. Similar to this, a look-up table for stationary operation and another for dynamic operation is suggested in [134]. These look-up tables are then weighted by a linear interpolation depending on the degree of dynamics. Another possibility is to add differential elements to the control output, which are again calibrated over the engine operation point [46].

### Control actuators

Depending on the engine configuration, there are various actuators possible to adjust the control variables. Common applied actuators are the exhaust gas recirculation valve  $s_{\text{egr}}$  and the guide vanes respectively the waste gate of the turbocharger  $s_{\text{t}}$ . Further actuators are the crank angle of main injection  $\varphi_{\text{mi}}$ , the throttle valve  $s_{\text{tv}}$ , the swirl flaps  $s_{\text{sf}}$ , the rail pressure  $p_{\text{rail}}$  and the low-pressure exhaust gas recirculation valve  $s_{\text{lp-egr}}$  if available. The number of utilised actuators should be identical to the number of control variables, and the selected control variables should not be redundant. An extraction of possible actuators and control variables is given in Tab. 5.3. The air to fuel ratio  $\lambda$  and the gas composition at outlet  $x_{\text{eng,out}}$  are equivalent by eq. (A.13). The rail pressure  $p_{\text{rail}}$  is again controlled by an inner cascade.

**Table 5.3:** Extraction of possible control variables and actuators for the control of a Diesel engine. Selected control variables should not be redundant and suit to the applied actuators.

Possible control variables		Possible actuators	
intake pressure	$p_{2i}$	turbocharger guide vanes	$s_{\text{t}}$
air mass per cycle	$m_{\text{air}}$	egr valve	$s_{\text{egr}}$
egr rate	$r_{\text{egr}}$	low-pressure egr valve	$s_{\text{egr}}$
egr mass per cycle	$m_{\text{egr}}$	crank angle of main injection	$\varphi_{\text{mi}}$
gas composition at intake	$x_{2i}$	throttle valve	$s_{\text{tv}}$
gas composition at outlet	$x_{\text{eng,out}}$	swirl flaps	$s_{\text{sf}}$
air to fuel ratio	$\lambda$	rail pressure	$p_{\text{rail}}$
NO <sub>x</sub> concentration	$c_{\text{nox}}$		
soot concentration	$c_{\text{soot}}$		

Common air path controls apply two control variables, mostly  $p_{2i}$  and  $m_{\text{air}}$  or  $r_{\text{egr}}$ , adjusted by the actuators  $s_{\text{egr}}$  and  $s_{\text{t}}$ . More actuators than control variables can be applied, if two or more actuators are coupled. An example is the coupling of the throttle valve  $s_{\text{tv}}$  with the egr valve  $s_{\text{egr}}$ . For small deviations, only the egr-valve is applied. If the egr valve is fully opened, the fresh air mass flow can further be reduced by closing the throttle valve. This can be necessary for negative transients in engine speed, as presented in Fig. 3.13. There, a fully opened egr valve does not reduce the fresh air mass flow sufficiently.

The performance of such a coupled actuator is shown in Fig. 5.2. The topmost plot shows the operation point trajectory and the dashed lines in the second and third plot the desired values for  $m_{\text{air}}$  and  $p_{2i}$ , which are the same as for the example in Fig. 3.13. The black lines in the lower two plots show the controller output with the actuators  $s_{\text{egr}}$  and  $s_t$  (index 1). There are significant deviations in  $m_{\text{air}}$  for negative transients in engine speed. These are avoided if the egr actuator  $s_{\text{egr}}$  is coupled with the throttle valve  $s_{\text{tv}}$ , as is shown by the grey lines in the lower two plots (index 2). The throttle valve is shown by the dashed lines. If the egr valve is fully closed, the throttle valve is applied to further reduce the air mass per cycle  $\dot{m}_{\text{air}}$ . Thus, only small deviations occur, as shown by the grey lines for  $m_{\text{air}}$  and  $p_{2i}$ .

### Controller designs

Besides the various control variables and actuators, there are several controller designs possible. PID controls with decouplings are presented in [9, 143], a model predictive control is shown in [140], an  $H_\infty$  approach in [73] and a flatness based approach in [84]. Overviews and comparisons of several controller designs are given in [175, 181].

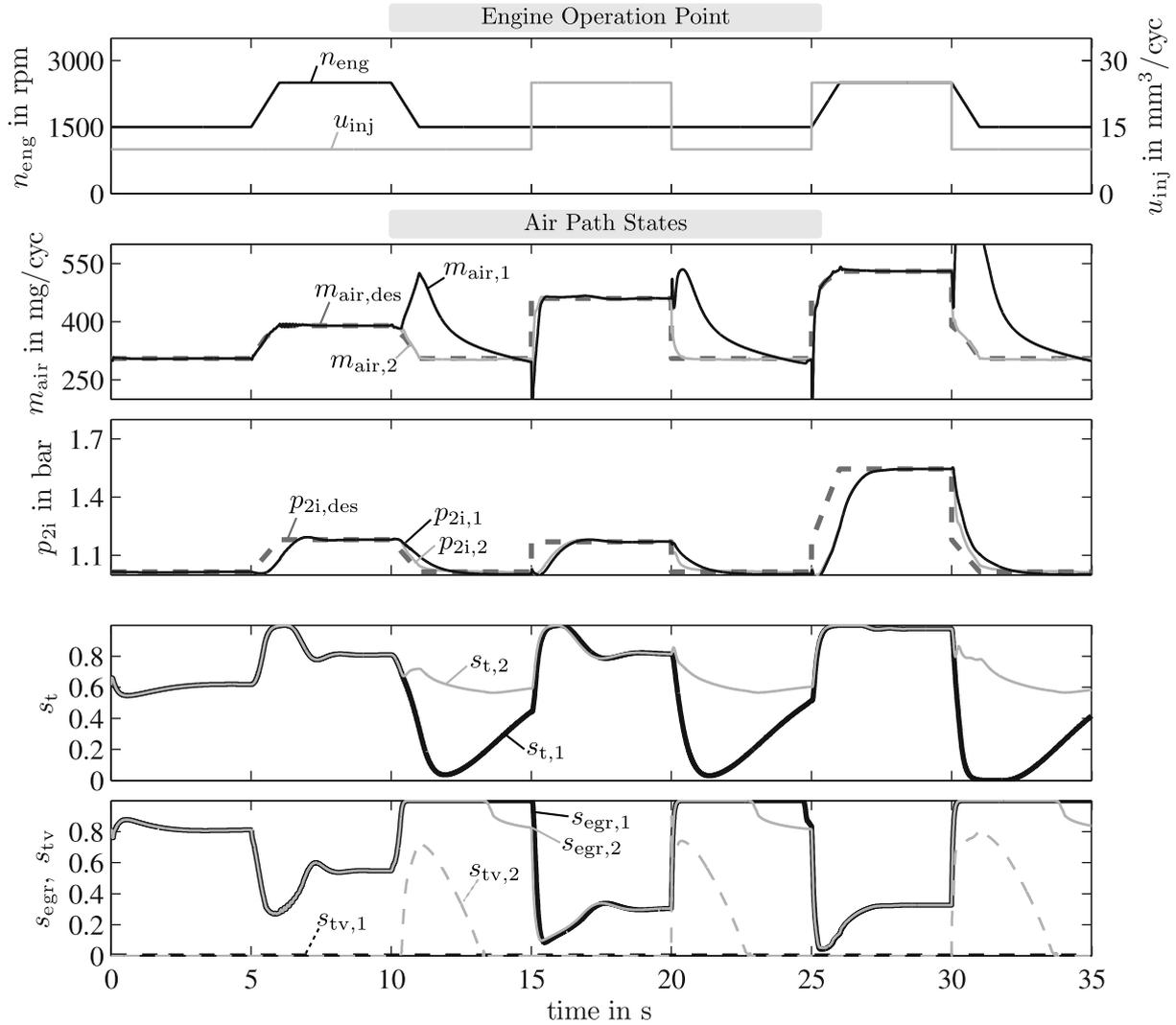
Control structures can be rated by a similar objective function as the one presented in eq. (5.1). The stationary calibrated values are then the desired values  $y_{\text{des}}$  and the differences to these are penalised. It is furthermore possible to penalise the actuating energy by an additional term  $\Delta \mathbf{s}(t)$  with a weighting factor  $\beta$ , see [63],

$$J_{\text{control}}(\mathbf{s}(t)) = \int_0^{T_{\text{end}}} (y(\mathbf{s}(t)) - y_{\text{des}}(t))^2 + \beta(\Delta \mathbf{s}(t))^2 dt. \quad (5.4)$$

This objective function is applied in [140, 143] for the design of model predictive controls of the air path and proposed for the design of PID controls in [181]. A general overview of tuning rules for PID controls is given in [10, 63]. For other control strategies, such as internal model control, the parameters are derived from the determination of a desired transfer function, without the explicit formulation of an objective function [63, 135].

### General engine control structure

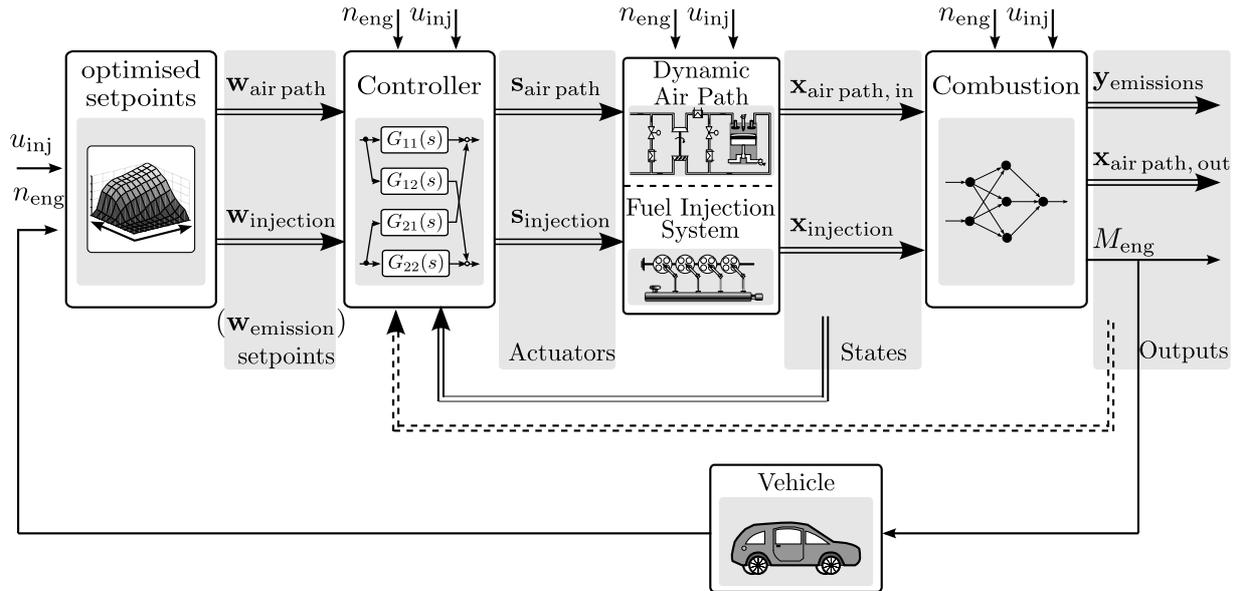
An overview of the general engine control structure is depicted in Fig. 5.3. The optimised set points depend on the engine operation point. These set points are the desired values for the applied control variables. Possible control variables are the states of the air path  $\mathbf{w}_{\text{air path}}$  and the injection system  $\mathbf{w}_{\text{injection}}$ . Alternative control variables are the engine emissions  $\mathbf{w}_{\text{emissions}}$ . The control variables are adjusted by the controller via the available actuators of the air path  $\mathbf{s}_{\text{air path}}$  and the injection system  $\mathbf{s}_{\text{injection}}$ . The resulting states of the dynamic air path  $\mathbf{x}_{\text{air path}}$  and the injection system  $\mathbf{x}_{\text{injection}}$  determine the outcome of the combustion process ( $\mathbf{y}_{\text{emissions}}$ ,  $\mathbf{x}_{\text{air path, out}}$  and  $M_{\text{eng}}$ ). Either the states of the air path and the injection system or the emissions can be feedback to the controller, depending on which control variables are applied. The individual blocks of this control structure are influenced by the engine operation  $n_{\text{eng}}$  and  $u_{\text{inj}}$ . The engine speed  $n_{\text{eng}}$  is determined by the states of the vehicle and the injection quantity  $u_{\text{inj}}$  results directly from the drivers request.



**Figure 5.2:** Typical transients in engine speed and injection quantity as presented in Fig. 3.13. Air path states  $m_{\text{air}}$  and  $p_{2i}$  are controlled by a decentralised PID control structure. Black lines show the results with actuators for the turbocharger  $s_t$  and the egr valve  $s_{\text{egr}}$  (index 1). Grey lines show the results with an actuator  $s_t$  and a coupled actuator applying  $s_{\text{egr}}$  and the throttle valve  $s_{\text{tv}}$  (index 2). The actuator trajectories are shown in the two lower plots by the black lines ( $s_t$  and  $s_{\text{egr}}$ ) and the grey lines ( $s_t$  and  $s_{\text{egr}}/s_{\text{tv}}$ ). The throttle valve is applied to further reduce the air mass flow.

## 5.4 Model-based smoke limitation

The optimisation of the trajectories of the air path actuators for the acceleration event presented in Sect. 5.2 showed that even with the knowledge of the engine operation point trajectory, the soot peak in the first few milliseconds can not be avoided. The fuel injection system responds considerably faster than the air path such that not enough fresh air is available for a sudden increase in injection quantity. The requested injection quantity can be realised instantaneously, whereas the required air mass flow rate follows with a significant dynamics, mainly depending on the dynamics



**Figure 5.3:** Block diagram of the general engine control structure. Desired values  $w$  for the air path states and the injection characteristics are given in dependence on the current engine operation point. The desired values are adjusted by the controller applying the available actuators  $s$ . The actuators act on the air path system and the injection system. The resulting states  $x$  influence the produced emissions and torque. Depending on the applied control variables either the engine states or the (simulated) emissions are feedback to the controller.

of the turbocharger and the egr system. Thus, for positive transients in injection quantity low air to fuel ratios can occur and cause high soot emissions. These soot emissions are visible as smoke and poorly affect the accumulated emissions.

While the air path states have not reached their stationary calibrated values, the last influence on the emissions can be realised by the injection system. One possibility to reduce the dynamic soot emissions, with small effects on the engine torque, is by calibrating set points for normal operation and smoke operation, as suggested in [134]. The regarded set points are the injection angle, the rail pressure and the recirculated exhaust gas. Even with these set points it is not possible to fully avoid soot peaks for major steps in injection quantity.

The last influence on the soot emissions is then given by a limitation of the injection quantity. This limitation can be determined by a limit value for the air to fuel ratio  $\lambda$  [124], translated into an injection quantity limit by eq. (A.5). The limitation of the injection quantity avoids the soot peak, but has a negative effect on the engine torque. Hence, such a limitation should not be too stringent. A limitation of the air to fuel ratio covers the main effect on the soot emissions, but there are further influences that are not regarded if a constant limit for  $\lambda$  is applied. To take the various influences on the soot emissions into account, the soot model from Chap. 3 is applied in the following for a *model-based smoke limitation*. The injection quantity is limited such that a predefined soot limit is not exceeded. Since the main influences are covered by the soot model, no additional calibration effort is necessary. This model-based smoke limitation can then again be simplified to an open

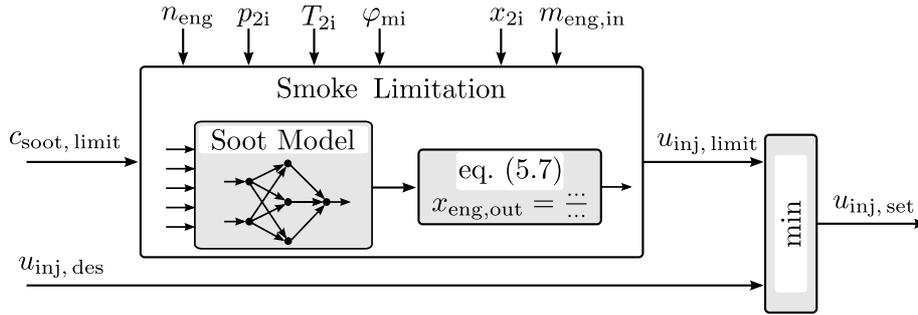
loop control structure, which has several advantages and is presented and discussed thereafter in Sect. 6.3.

### 5.4.1 Functional structure of the smoke limitation

The block diagram of the model-based smoke limitation is depicted in Fig. 5.4. The desired injection quantity  $u_{inj,des}$  is limited to  $u_{inj,limit}$  with utilisation of the soot model, the current engine speed, the current air path states and the current injection characteristics. The applied soot model, as presented in Chap. 3, can be written as

$$c_{soot}(n_{eng}, u_{inj}, x_{eng,out}, p_{2i}, \varphi_{mi}, T_{2i}) = \sum_{j=1}^M \Phi_j(n_{eng}, u_{inj}) c_{soot,j}(x_{eng,out}, p_{2i}, \varphi_{mi}, T_{2i}). \quad (5.5)$$

Note that the model for the soot concentration  $c_{soot}$  is applied rather than the model for the soot mass flow rate  $\dot{m}_{soot}$ , since the visibility of soot is described by its concentration. If the soot model is extended by further model inputs, the additional inputs can be included analogously in eq. (5.5).



**Figure 5.4:** Block diagram of the model-based smoke limitation. The desired injection quantity  $u_{inj,des}$  is limited to  $u_{inj,limit}$  such that a predefined soot limit  $c_{soot,limit}$  is not exceeded. The soot model is applied with inputs of the air path states, the engine speed and the injection characteristics. Together with the formula for the gas composition after combustion, eq. (5.8), an inversion of this soot model is performed with regard to  $u_{inj}$ .

To determine the limiting value for the injection quantity  $u_{inj,limit}$ , the soot model is inverted with regard to  $u_{inj}$ ,

$$u_{inj,limit} = c_{soot}^{-1}(c_{soot,limit})|_{n_{eng}, x_{eng,out}, p_{2i}, \varphi_{mi}, T_{2i}} \quad (5.6)$$

with the predefined soot limit

$$c_{soot,limit} = 15 \frac{\text{mg}}{\text{m}^3}. \quad (5.7)$$

Problematic for this inversion is the dependence of the local models  $c_{soot,j}$  on the the gas composition after combustion  $x_{eng,out}$ , which again depends on the injection quantity  $u_{inj}$ ,

$$x_{eng,out} = \frac{x_{2i} m_{eng,in} - u_{inj} \rho_{Diesel} L_{st}}{m_{eng,in} + u_{inj} \rho_{Diesel}}. \quad (5.8)$$

The total gas mass  $m_{\text{eng,in}}$  and the gas composition at intake  $x_{2i}$  are air path states, see App. A. The stoichiometric air to fuel ratio  $L_{\text{st}}$  and the density of Diesel  $\rho_{\text{Diesel}}$  are constant fuel characteristics.

Given the soot model as in eq. (5.5), the injection quantity  $u_{\text{inj}}$  appears in the weighting function  $\Phi_j$  and it appears in the local models  $c_{\text{soot},j}$  by the gas composition after combustion  $x_{\text{eng,out}}$ . It is introduced in the nominator and denominator of eq. (5.8). This complicates an analytic inversion of the soot model with regard to  $u_{\text{inj}}$ , why a numerical inversion is presented in the following. The individual steps are described, to evaluate the computational effort for an implementation in the ECU.

### 5.4.2 Implementation in the ECU

For an inversion of the soot model, at first all constant values are inserted in eq. (5.5). Regarding the smoke limitation, these are all engine states besides the injection quantity. Hence, the variables  $n_{\text{eng}}$ ,  $x_{2i}$ ,  $m_{\text{eng,in}}$ ,  $p_{2i}$ ,  $\varphi_{\text{mi}}$  and  $T_{2i}$  are inserted such that the soot model in eq. (5.5) simplifies to

$$c_{\text{soot}}(u_{\text{inj}}) = \sum_{j=1}^{M_{u_{\text{inj}}}} \Phi_j(u_{\text{inj}}) c_{\text{soot},j}(u_{\text{inj}}). \quad (5.9)$$

The local models  $c_{\text{soot},j}$  are in Chap. 3 given as polynomials of order three. Therefore, eq. (5.9) can also be written as

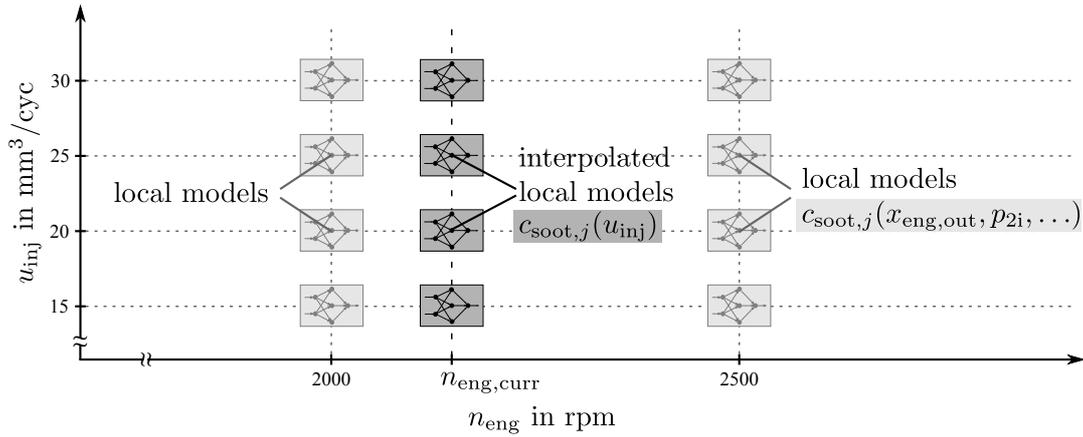
$$c_{\text{soot}}(u_{\text{inj}}) = \sum_{j=1}^{M_{u_{\text{inj}}}} \Phi_j(u_{\text{inj}}) (a_j x_{\text{eng,out}}^3(u_{\text{inj}}) + b_j x_{\text{eng,out}}^2(u_{\text{inj}}) + c_j x_{\text{eng,out}}(u_{\text{inj}}) + d_j). \quad (5.10)$$

with  $x_{\text{eng,out}}$  as in eq. (5.8). Since  $n_{\text{eng}}$  is regarded as constant,  $\Phi_j$  simplifies to a linear interpolation. The coefficients  $a_j$  to  $d_j$  are derived from inserting the engine states into the local models  $c_{\text{soot},j}$  in eq. (5.5) and can in general be written as

$$a_j, b_j, c_j, d_j = f_j(n_{\text{eng}}, p_{2i}, \varphi_{\text{mi}}, T_{2i}). \quad (5.11)$$

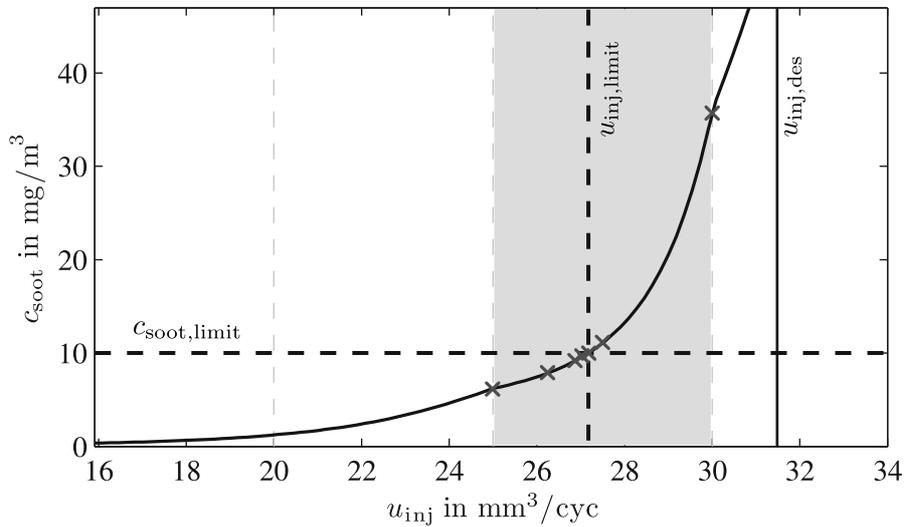
This simplification to eq. (5.9) is also depicted in Fig. 5.5. The light grey models show the local models of the applied global-local soot model,  $c_{\text{soot},j}(x_{\text{eng,out}}, p_{2i}, \varphi_{\text{mi}}, T_{2i})$ . These are interpolated with regard to the current engine speed  $n_{\text{eng,curr}}$  and after insertion of all constant engine states, the dark grey local models result,  $c_{\text{soot},j}(u_{\text{inj}})$ . For these models, the predicted soot depends only on the injection quantity, see also eq. (5.9). This requires some computing time, but the soot model can thus easily be evaluated for several injection quantities.

At first, the soot model is evaluated for the desired injection quantity  $u_{\text{inj,des}}$ . If the limit is not exceeded, the injection quantity need not to be limited. If the predicted soot exceeds the limit, the next lower local model is evaluated at the applied resolution of  $\Delta u_{\text{inj}} = 5 \text{ mm}^3/\text{cyc}$ , see the cross in Fig. 5.6 at  $u_{\text{inj}} = 30 \text{ mm}^3/\text{cyc}$ . If the soot limit is still exceeded, the next lower local model is evaluated and so on, until the soot limit is no longer violated. Then the maximum injection quantity



**Figure 5.5:** Simplification of the global-local soot model for the smoke limitation. All engine states besides the injection quantity are regarded as constants and inserted into the local soot models. The local models (light grey) simplify by a linear interpolation with regard to the current engine speed  $n_{eng,curr}$  to local models  $c_{soot,j}(u_{inj})$  (dark grey). These local models depend only on the injection quantity  $u_{inj}$ , since all other inputs are regarded as constant.

is in the grey shaded region, see Fig. 5.6. With five more evaluations inside the grey shaded region (marked as crosses), the accuracy of the calculated inverse can be determined up to an uncertainty of  $0.1 \text{ mm}^3/\text{cyc}$ . This is sufficient for a smoke limitation and appropriate with regard to model uncertainties.



**Figure 5.6:** Determination of the limit value for the injection quantity  $u_{inj,limit}$  with regard to a predefined soot limit  $c_{soot,limit}$ . If the soot limit is exceeded for the desired injection quantity  $u_{inj,des}$ , the injection quantity is reduced until the limiting value is found. Several evaluations (marked as crosses) are necessary for the presented numerical inversion.

Hence, the determination of the limit value  $u_{inj,limit}$  by the model-based smoke limitation requires at first an insertion of the current engine states in the soot model. The high dimensional global-local

model structure in eq. (5.5) then simplifies to a global-local model structure in  $u_{inj}$ , see eq. (5.9). The inverse is then determined by several evaluations of this model, as depicted in Fig. 5.6.

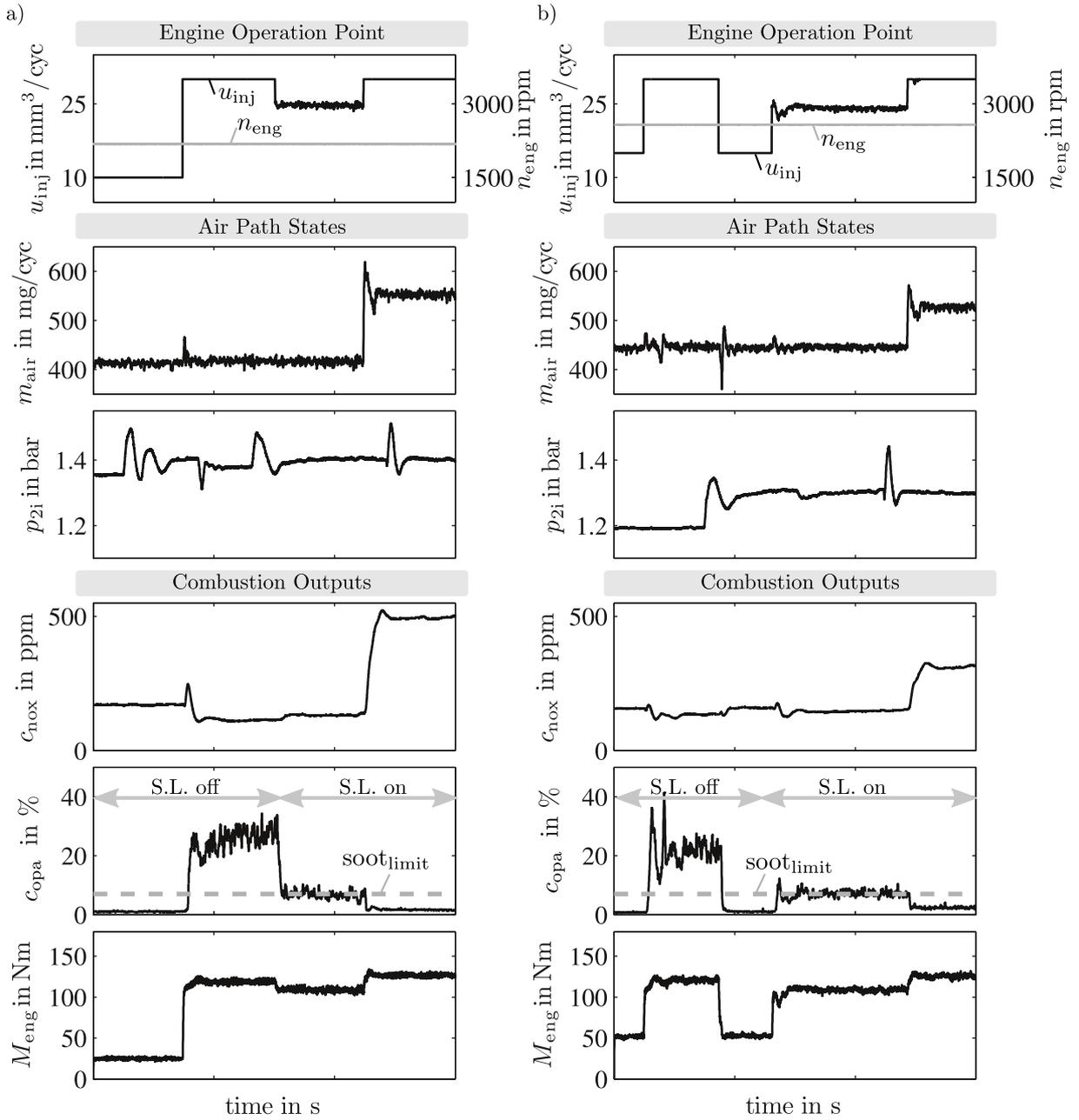
The model-based smoke limitation has the advantage that no further calibration effort is necessary. If the soot model is given as in Chap. 3, the smoke limitation can directly be applied. A drawback is that the parameters of the global-local polynomial model structure are difficult to interpret, why a fine tuning of such a model-based smoke limitation is a complicated task. If a fine tuning is desired, an open loop control with look-up tables is favoured. The derivation of such an open loop control from the presented model-based smoke limitation is presented in Sect. 6.3 in simplified form. In the following, test bed results are shown for the previously presented model-based smoke limitation.

### 5.4.3 Test bed results

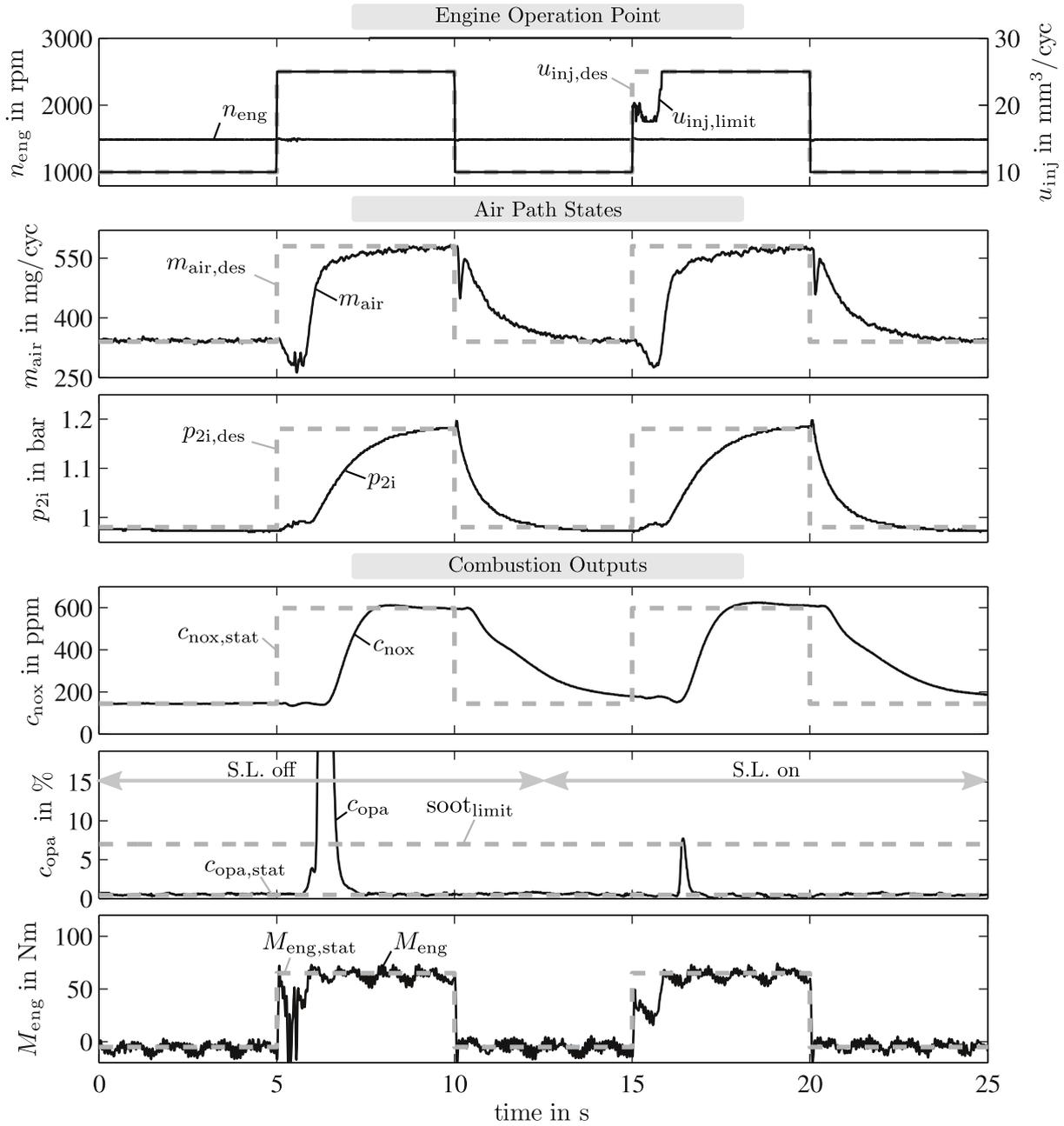
A smoke limitation is only necessary for engine states with low air to fuel ratios. These states are usually avoided by the engine control. To illustrate the performance of the presented model-based smoke limitation, the engine is intentionally operated at a stationary state with a low air to fuel ratio. Therefore, the egr valve is kept open during a step in injection quantity. Due to the recirculated exhaust gas, not enough fresh air enters the combustion chamber. If then the injection quantity is increased, the predefined soot limit is violated, see opacity plot in Fig. 5.7 a). Opacity highly correlates to the soot measurement and is applied here since the Micro Soot sensor is not available for this measurement ( $15 \text{ mg/m}^3 \approx 7\%$ ). For the first 30 s the smoke limitation is switched off (S.L. off), why the predefined limit is exceeded. Switching the smoke limitation on at  $t = 30 \text{ s}$  (S.L. on), the injection quantity is limited such that the soot does not exceed the predefined limit. At  $t = 45 \text{ s}$  the egr valve is closed such that additional fresh air can enter the combustion chamber. The smoke limitation does no longer detect a violation of the soot limit and releases the injection quantity to its desired value.

In Fig. 5.7 b) two steps in injection quantity are performed. For the first step the smoke limitation is switched off (S.L. off) and for the second step it is switched on (S.L. on). Again the egr valve is kept open such that not enough fresh air is available. The soot limit is therefore again exceeded for the step with the smoke limitation switched off. The injection quantity is then reduced again and the smoke limitation is switched on. The same step is performed now with the smoke limitation switched on. The injection quantity is limited such that the predefined soot limit is not exceeded. The intake pressure  $p_{2i}$  oscillates during this step, which also influences the formation of soot. The model-based smoke limitation reacts on these oscillations by adapting the injection quantity accordingly, see  $u_{inj}$  at  $t = 125 \text{ s}$ . At  $t = 150 \text{ s}$  the egr valve is closed again and the smoke limitation releases the desired injection quantity.

In Fig. 5.8 again two steps in injection quantity are performed, one with the smoke limitation switched off and the other with the smoke limitation switched on. In this test, the dynamic performance of the smoke limitation is investigated for a typical step in injection quantity. Air path states are calibrated for the operation points before the step and the target operation point. The air



**Figure 5.7:** Steps in injection quantity with the smoke limitation switched off and on. a) Step in injection quantity  $u_{inj} = 10 \rightarrow 30 \text{ mm}^3/\text{cyc}$  with  $n_{eng} = 2200 \text{ rpm}$ .  $p_{2i} = 1.3 \text{ bar}$  and  $\varphi_{Q50} = 14^\circ\text{CA}$  are controlled at their set points. The smoke limitation is switched on at  $t = 30 \text{ s}$ . b) Two steps in injection quantity  $u_{inj} = 15 \rightarrow 30 \text{ mm}^3/\text{cyc}$  with  $n_{eng} = 2600 \text{ rpm}$  with the smoke limitation switched off for the first step and switched on for the second step.  $p_{2i} = 1.4 \text{ bar}$  and  $\varphi_{Q50} = 14^\circ\text{CA}$  are controlled at their set points.



**Figure 5.8:** Step in injection quantity  $u_{inj} = 10 \rightarrow 25 \text{ mm}^3/\text{cyc}$  with  $n_{eng} = 1500 \text{ rpm}$ . Air path actuators are controlled by open loop controls. Desired values are plotted as grey dashed lines.  $\varphi_{Q50} = 15^\circ\text{CA}$  is controlled by a closed loop control. The smoke limitation is switched on at  $t = 12.5 \text{ s}$ .

path actuators  $s_{egr}$  and  $s_t$  are controlled by open loop controls. Because of these open loop controls the air path states require a significant time to reach steady state. During this time, not enough fresh air is available such that a soot peak arises, see opacity peak at  $t = 6.1 \text{ s}$ . Note that the delay of the soot peak is due to the measurement dynamics. The soot peak exceeds the predefined limit for about 0.6 s. After this, enough fresh air is available and the soot emissions are reduced.

For the second step, starting at  $t = 15$  s, the smoke limitation is switched on. The model-based smoke limitation directly detects a violation of the soot limit and lowers the injection accordingly, see  $u_{inj}$  in topmost plot. When the fresh air mass flow increases again, the limit for  $u_{inj}$  is released. The soot peak is significantly reduced and does not exceed the predefined limit. Again, the measurement dynamics can well be seen. Limiting the injection quantity also limits the engine torque. It can however be seen that the engine torque is also lowered similarly for the step with the smoke limitation switched off. This reduced engine torque for the not limited case is because of an incomplete combustion, which is due to the low air to fuel ratio and therefore due to the lack of fresh air required for a complete combustion. Hence, the torque is only little effected by such a smoke limitation but the soot emissions are significantly reduced.

## 5.5 Summary

The optimisation of the dynamic engine operation is regarded. An objective function to rate the dynamic behaviour is presented, which penalises the deviations of the dynamic emissions from the stationary calibrated values. Based on this objective function, a typical acceleration event is optimised. The regarded optimisation variables are the discretised trajectories of the turbocharger guide vanes  $s_t$  and the egr valve  $s_{egr}$ . The optimisation results are superior to an open loop control of the actuators and also superior to a decentralised PID control of the air path states. The computing time of the optimisation is relatively high, why such an optimisation can not be calculated online on the ECU during engine operation. The results are however a benchmark for the design of closed loop controls and can be utilised to design feedforward controls.

An overview of various closed loop controls for dynamic engine operation is given. The task of a closed loop control is to accelerate the dynamics of the air path system such that the dynamic emissions are reduced. Common approaches control the emissions indirectly by controlling the air path states. Alternative approaches control the emissions directly using the measured emissions. Since emission sensors are affected by measurement dynamics and since these sensors are cost and maintenance intensive, virtual sensors are applied. Virtual emission sensors are based on emission models and related measurements, such as the air to fuel ratio that relates to the soot emissions. These and other possible control variables are listed together with possible control actuators. A general engine control structure is presented.

Even with a closed loop control, low air to fuel ratios can occur during transients, since the dynamics of the injection system are considerably faster than the dynamics of the air path. Low air to fuel ratios cause high soot emissions, which are visible as smoke and poorly affect the accumulated emissions. To avoid these emissions, a model-based smoke limitation is presented, which limits the injection quantity. Such a limitation should not be too stringent, as it also reduces the engine torque. Therefore, all influences on the soot emissions need to be regarded, why the soot model, as presented in Chap. 3, is applied. Since an analytical inversion of the global-local soot model is highly complicated, the soot model is inverted numerically. Results from the engine test

---

bed with the smoke limitation switched off and on are shown. Due to the utilisation of the soot model, no additional calibration effort is necessary. A fine tuning of such a structure is however complicated, why an open loop control of the smoke limitation is derived from this model-based smoke limitation in Sect. 6.3.

---

## 6 Comparison of Model Structures for Control Applications in the ECU

---

The suitability of various model structures for an implementation in the ECU is investigated in this chapter for three different control applications. At first, the model structures presented in Chap. 2 are compared to each other with regard to a  $\text{NO}_x$  emission model. Applied evaluation criteria are the model accuracy, the computing time and the required memory. Thereafter, a global look-up table model of the engine torque is presented. The invertibility of this structure is discussed for a control of the injection quantity. Finally, an open loop control of the smoke limitation is derived from the model-based smoke limitation, presented in Sect. 5.4. Conclusions for the usability of look-up tables and alternative model structures for an implementation in the ECU are given.

---

Complex control functions are necessary for an accurate engine control. The number and the complexity of these control functions increased steadily in recent years. Therefore, the computing and memory capacity of ECUs increased simultaneously. This enables modern ECUs to calculate complex model-based control functions, but they are always limited to real time capability. Depending on the field of application, the available computing time for a control function ranges from a few milliseconds to several seconds. The former is for example given for the control of injection characteristics and the latter for the heating-up period after the engine start. Hence, the suitability of a control function depends on the regarded process part and the related requirements.

In the following, three different control functions are regarded. In Sect. 6.1 the simulation of a  $\text{NO}_x$  model on the ECU is investigated. Therefore, the model structures presented in Chap. 2 are compared to each other with regard to an implementation in the ECU. The accuracy, the computing time and the required memory are determined for each of these model structures. Then in Sect. 6.2, the implementation of a global torque model is presented. Since the torque model needs to be invertible in the ECU, a look-up table structure is applied. In Sect. 6.3 an open loop control of the smoke limitation is derived from the model-based smoke limitation presented in Sect. 5.4. A look-up table structure is applied, which has the advantage of interpretable parameters. Thus, the functionality of the smoke limitation can easily be ensured and the derived results can intuitively be fine tuned on the engine test bed.

### 6.1 Comparison of model structures for a $\text{NO}_x$ model

The efficient storage and the fast computation of model-based approaches in the ECU is a challenging task. Two-dimensional look-up tables are mostly applied in modern ECUs because of their

fast computation. Higher dimensional relations are then mapped by a combination of several two-dimensional look-up tables, see Sect. 2.1.3. The determination of a suited look-up table structure is a time-consuming task and the number of utilised two-dimensional look-up tables may be large for higher dimensional relations. To evaluate the usability of alternative model structures for an implementation in the ECU, the model structures introduced in Chap. 2 are compared to each other with regard to the implementation of a NO<sub>x</sub> emission model. The considered model structures are look-up tables, LOLIMOT, LOPOMOT and the Kernel model.

To model the NO<sub>x</sub> emissions, the global-local model structure, as introduced in eq. (3.18) in Sect. 3.2, is applied for all approaches. Hence, the global structure is the same for the various approaches, but they differ in the structure of the local models,  $\dot{m}_{\text{nox},j}(\mathbf{u})$ . There are  $M = 21$  local models to be identified, distributed over the engine operation points, as depicted in Fig. 3.4. The weighting of the local models is again given by a bilinear interpolation  $\Phi_j$  such that the general formulation is analogue to eq. (3.18),

$$\dot{m}_{\text{nox}}(\mathbf{z}, \mathbf{u}) = \sum_{j=1}^M \Phi_j(\mathbf{z}) \dot{m}_{\text{nox},j}(\mathbf{u}). \quad (6.1)$$

Given this bilinear interpolation, at most four local models are relevant for a function evaluation, which reduces the required computing time for all approaches. The inputs of the local models are, as in Sect. 3.1, the air mass per cycle  $m_{\text{air}}$ , the intake pressure  $p_{2i}$ , the crank angle of 50% mass fraction burnt  $\varphi_{Q50}$  and the intake temperature  $T_{2i}$ ,

$$\mathbf{u}^T = [m_{\text{air}}, p_{2i}, \varphi_{Q50}, T_{2i}]. \quad (6.2)$$

The inputs of the weighting function are given by the engine operation point, defined as

$$\mathbf{z}^T = [n_{\text{eng}}, u_{\text{inj}}]. \quad (6.3)$$

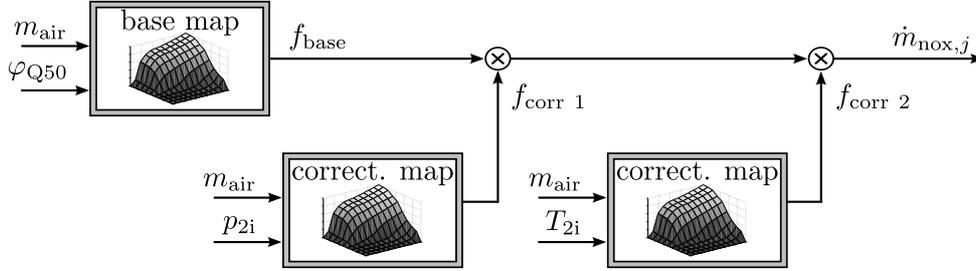
Such a structure can for look-up tables be interpreted as nested model structure, see Fig. 2.6. For the LOPOMOT model, no further partitions are applied for the local models such that they are identical to adaptive polynomials. Descriptions of the various local model structures are given in Sect. 6.1.1 together with a comparison of the achieved model qualities. In Sect. 6.1.2 the effort for a single model evaluation is regarded. This is important for an implementation in the ECU, since the computing capacity and the computing time are limited. Besides the computing capacity, also the memory on an ECU is limited, why this is investigated in Sect. 6.1.3. Finally, conclusions about the usability of the compared model structures are given in Sect. 6.1.4.

### 6.1.1 Model accuracy

The general model structures are introduced in Chap. 2. In the following the particular realisations for the local models are shortly described and the model qualities are stated. The applied criteria of fit to compare the models are the coefficient of determination  $R^2$  and the root mean square error RMSE. These are determined on the training and validation data presented in Sect. 3.3.

### Look-up tables

The local models  $\dot{m}_{\text{nox},j}(\mathbf{u})$  depend on four input variables, see eq. (6.2), whereas the regarded look-up tables are limited to two inputs. To map this higher dimensional relation, a combination of several two-dimensional look-up tables is applied. The utilised look-up table structure is derived from trial and error and depicted in Fig. 6.1.



**Figure 6.1:** Multiplicative local look-up table structure to model  $\dot{m}_{\text{nox},j}(\mathbf{u})$ . Inputs of the look-up tables and the multiplicative composition of the two-dimensional look-up tables are derived from trial and error. This look-up table structure is applied for each of the  $M = 21$  engine operation points. The global model output is determined by eq. (6.1).

To identify this look-up table structure, at first the base look-up table is determined by a regression. The applied inputs  $m_{\text{air}}$  and  $\varphi_{Q50}$  have the highest contribution for such a regression, why these are selected for the base look-up table. Thus, the four dimensional relation is approximated by a two-dimensional look-up table. Then the residuals are computed and the best performing correction look-up table is selected by trial and error. A multiplicative correction look-up table shows the best improvement, applying the inputs  $m_{\text{air}}$  and  $p_{2i}$ . Again, a regression with regard to these inputs and the regarded residuals is performed to identify the correction look-up table. Then, a further correction look-up table is identified analogously with the selected inputs  $m_{\text{air}}$  and  $T_{2i}$ . Hence, the local model can be written as

$$\dot{m}_{\text{nox},j}(\mathbf{u}) = f_{\text{base}}(m_{\text{air}}, \varphi_{Q50}) \cdot f_{\text{corr } 1}(m_{\text{air}}, p_{2i}) \cdot f_{\text{corr } 2}(m_{\text{air}}, T_{2i}). \quad (6.4)$$

The look-up table inputs are each discretised in ten equidistant steps such that the discretisation is given by

$$\Delta \mathbf{u}^T = [68.8 \frac{\text{mg}}{\text{cyc}}, 0.11 \text{ bar}, 1.36 \text{ }^\circ\text{CA}, 13.6 \text{ }^\circ\text{C}]. \quad (6.5)$$

The distribution of grid points is the same for all  $M = 21$  local models, each defined for a fixed operation point. Since the drivability space varies with the engine operation point, there are regions in each look-up table that are not covered by measurements. Therefore, the regularisation as introduced in Sect. 2.1.2 is applied.

The attained model qualities are  $R^2 = 0.991$  and  $\text{RMSE} = 2.08 \text{ mg/s}$  for the training data and  $R^2 = 0.982$  and  $\text{RMSE} = 2.56 \text{ mg/s}$  for the validation data.

## LOLIMOT

The LOLIMOT model approximates each local model  $\dot{m}_{\text{nox},j}(\mathbf{u})$  by a superposition of affine functions. The partitions on which the local affine models are valid are determined automatically by the tree construction algorithm. This algorithm divides the local drivability space by axis orthogonal splits in the input  $\mathbf{u}$ . The maximum number of local model partitions is limited to ten and the best number of model partitions is selected by a heuristic approximation of Akaike's Information Criteria (AIC) implemented in the algorithm, see [122]. The local LOLIMOT model with  $p$  inputs and  $M_L$  partitions is then given by

$$\dot{m}_{\text{nox},j}(\mathbf{u}) = \sum_{i=1}^{M_L} \Phi_{\text{LOLIMOT},i}(\mathbf{u}) \cdot (w_{0,i} + w_{1,i}u_1 + \dots + w_{p,i}u_p) \quad (6.6)$$

For the 21 local operation points, there are in average  $M_L = 8.38$  model partitions.

The attained model qualities are  $R^2 = 0.992$  and  $\text{RMSE} = 1.93$  mg/s for the training data  $R^2 = 0.985$  and  $\text{RMSE} = 2.38$  mg/s for the validation data.

## LOPOMOT

Since LOPOMOT utilises local polynomial models of order three, there are no further partitions of the input space necessary. Regarding the local models, the LOPOMOT model simplifies to an adaptive polynomial model, as described in Sect. 2.2,

$$\dot{m}_{\text{nox},j}(\mathbf{u}) = w_1x_1 + w_2x_2 + \dots + w_nx_n \quad (6.7)$$

with 
$$x_i \in \mathcal{A} = \{1, u_1, u_2, \dots, u_4, u_1^2, u_1u_2, \dots, u_4^3\}. \quad (6.8)$$

This model is the same as the one presented in Chap. 3 for  $\dot{m}_{\text{nox}}$ .

The attained model qualities are  $R^2 = 0.996$  and  $\text{RMSE} = 1.45$  mg/s for training data and  $R^2 = 0.988$  and  $\text{RMSE} = 2.09$  mg/s for validation data.

## Kernel model

The Kernel model is a data driven approach, described in detail in Sect. 2.5. It is a class of model structures from which the applied type utilises one Kernel function for each measurement point ( $\mathbf{u}(i), y(i)$ ) of the training data. Therefore, the formulation of a local model is given by

$$\dot{m}_{\text{nox},j}(\mathbf{u}) = \frac{\sum_{i=1}^{M_L} K_{\mathbf{h}}(\mathbf{u}, \mathbf{u}(i))y(i)}{\sum_{i=1}^{M_L} K_{\mathbf{h}}(\mathbf{u}, \mathbf{u}(i))}. \quad (6.9)$$

with  $M_L$  being the number of local measurements. The parameters to be identified are the elements of the bandwidth vector  $\mathbf{h}$  of the kernel function  $K_{\mathbf{h}}$ . The kernel function is a Gaussian, as presented

in eq. (2.67), and there is one bandwidth parameter for each dimension of  $\mathbf{u}$ . To identify the bandwidths, the ten-fold cross-validation error on the training data is minimised by applying a numerical optimisation algorithm, see Sect. 4.2.1. The bandwidths are individual identified for each of the 21 local models. The average bandwidth parameters over the 21 operation points are

$$\bar{\mathbf{h}} = \left[ 17.4 \frac{\text{mg}}{\text{cyc}}, 0.05 \text{ bar}, 1.92 \text{ }^\circ\text{CA}, 28.2 \text{ }^\circ\text{C} \right]. \quad (6.10)$$

There are various alternative Kernel models to the one presented in Sect. 2.5, which differ in type and number of applied kernel function. The various approaches possess different properties, why the regarded Kernel model shall be regarded as an example for the class of Kernel models.

The attained model qualities are  $R^2 = 0.997$  and  $\text{RMSE} = 1.24 \text{ mg/s}$  for training data and  $R^2 = 0.982$  and  $\text{RMSE} = 2.58 \text{ mg/s}$  for validation data.

## Comparison

The qualities of the various model structures are summarised in Tab. 6.1. The results show that all models attain an  $R^2$  above 0.99 on training data and are therefore all able to model the measured  $\text{NO}_x$  emissions with minor differences in accuracy. The validation error on the other hand shows some drawbacks for the Kernel model and the look-up table structure.

**Table 6.1:** Model qualities of the four regarded model structures for models of  $\dot{m}_{\text{nox}}(\mathbf{z}, \mathbf{u})$ . Model qualities are stated in coefficient of determination  $R^2$  and root mean squared error RMSE for the training and validation data as presented in Sect. 3.3.

Model	Training		Validation	
	$R^2$	RMSE	$R^2$	RMSE
Look-up tables	0.991	2.08 mg/s	0.982	2.56 mg/s
LOLIMOT	0.992	1.93 mg/s	0.985	2.38 mg/s
LOPOMOT	0.996	1.45 mg/s	0.988	2.09 mg/s
Kernel model	0.997	1.24 mg/s	0.982	2.58 mg/s

The LOPOMOT structure shows the best performance, which is due to its possibility to adapt to the non-linearity of the process by selecting the significant regressors for a local model. The LOLIMOT structure also adapts to the process by partitions of the input space, why this structure results in comparable results. The look-up table structure suffers from its equidistant distribution of grid points. Since for each local model the same discretisation is applied, some regions of the look-up tables are not covered with measurements. The quality of the Kernel model mainly depends on the coverage of the input space. Since the training values are part of the model structure, it shows the best performance on training data. The quality on the validation data depends on the distances to the training values and is the worst for the regarded model structures.

### 6.1.2 Computing time

To compare the computational effort for a model simulation, the required floating point operations (FLOPs) are counted for each model structure. It is distinguished between multiplications '\*', additions '+' and evaluations of an exponential function 'exp'.

In the following, the required FLOPs for an evaluation of one local model  $\dot{m}_{\text{nox},j}(\mathbf{u})$  are regarded. Besides these FLOPs, there are additional FLOPs necessary to determine the global model output  $\dot{m}_{\text{nox}}(\mathbf{z}, \mathbf{u})$ , see eq. (6.1) and eq. (3.18). To compare the four model structures, the FLOPs required for one simulation of the global  $\dot{m}_{\text{nox}}(\mathbf{z}, \mathbf{u})$  are then summarised and compared.

#### Look-up tables

The required steps for an evaluation of a two-dimensional look-up table are reviewed in the following. At first, the four grid points in the lattice structure surrounding the input  $\mathbf{u}$  need to be determined. Depending on the applied grid point distributions, several comparisons might be necessary. Since there are efficient algorithms to find the surrounding grid points, see [145], these operations are not considered in the following.

Once the four surrounding grid points are determined, the bilinear interpolation is computed. To determine the weights for the bilinear interpolation  $\Phi_i(\mathbf{u})$ , see eq. (2.1), the area interpolation is applied, see Fig. 2.1 b). Therefore, eq. (2.3) is evaluated for all four surrounding grid points. An evaluation requires two additions and two multiplications for each grid point, which makes eight additions and eight multiplications for one look-up table. The multiplication in the denominator in eq. (2.3b) is saved, since the area of a grid element is a constant. Given the weights for the grid points, the look-up table output is determined by multiplying the weights with the corresponding grid point heights and summing them up, see eq. (2.1). Therefore, another four multiplications and three additions are necessary. The required FLOPs for an evaluation of a two-dimensional look-up table are therefore

$$\begin{aligned}
 + & : 8 + 3 \\
 * & : 8 + 4 \\
 \text{exp} & : -
 \end{aligned} \tag{6.11}$$

Since the look-up table structure presented in Fig. 6.1 consists of three two-dimensional look-tables, three of these evaluations and two more multiplications are required to calculate the output for one local model  $\dot{m}_{\text{nox},j}(\mathbf{u})$ .

#### LOLIMOT

LOLIMOT adapts to the systems non-linearities by axis orthogonal partitions in the input space. In each partition an affine model is identified and the partitions are automatically determined by the tree construction algorithm, see Fig. 2.11. Depending on the non-linearity of the system, several

partitions might be necessary for a satisfying model quality. The number of model partitions directly influences the computing time.

The output of a local LOLIMOT model with  $p$  inputs and  $M_L$  partitions is given by eq. (6.6). An evaluation of eq. (6.6) requires for each of the  $M_L$  partitions, one multiplication by the weighting function,  $p$  multiplications of the inputs by their parameters and  $p$  additions inside the brackets. Furthermore, there are  $M_L - 1$  additions due to the sum over the  $M_L$  partitions. Thus, the required FLOPs for an evaluation of eq. (6.6) are

$$\begin{aligned}
 + &: M_L \cdot (p + 1) - 1 \\
 * &: M_L \cdot (p + 1) \\
 \mathbf{exp} &: -
 \end{aligned} \tag{6.12}$$

Furthermore, there are additional FLOPs necessary for the evaluation of the  $M_L$  weighting functions,  $\Phi_{\text{LOLIMOT},i}(\mathbf{u})$ . For the Gaussian weighting function, see eq. (2.48), these additional FLOPs are

$$\begin{aligned}
 + &: 2p \cdot M_L + 1 \\
 * &: (2p + 1) \cdot M_L + 1 \\
 \mathbf{exp} &: M_L
 \end{aligned} \tag{6.13}$$

The regarded  $\text{NO}_x$  model has  $p = 4$  local model inputs and in average  $M_L = 8.38$  partitions for a local model. For an implementation in the ECU, especially the evaluations of the Gaussian are critical. These evaluations can be avoided if alternative weighting functions, such as polynomial approximations of the Gaussian [119], are applied.

## LOPOMOT

Due to the increased local model complexity of LOPOMOT, there are no further partitions necessary to model  $\dot{m}_{\text{nox},j}(\mathbf{u})$ . Thus, the local model simplifies to a local polynomial, see eq. (6.7), and the computational effort only depends on the number of local regressors  $n$ . There are less than  $n$  multiplications necessary to compute the regressors from the inputs. Linear regressors require no multiplication, regressors such as  $u_1 u_2$  require one multiplication and regressors such as  $u_1^2 u_2$  are computed by multiplying  $u_1 u_2$  by  $u_1$ . Given the regressors, there are another  $n - 1$  multiplications necessary to multiply the regressors by their coefficients (the offset needs no multiplication). Finally the terms are summed up to the local model output such that the required FLOPS are

$$\begin{aligned}
 + &: \leq n \\
 * &: \leq n + n - 1 \\
 \mathbf{exp} &: -
 \end{aligned} \tag{6.14}$$

The number of FLOPs are given as an upper bound.  $n$  is the number of potential regressors and there are in general less than  $n$  regressors selected for a local model, which further reduces the

computing time. The number of potential regressors  $n$  depends on the selected polynomial order and the dimension of the input  $\mathbf{u}$ . For the regarded emission model, an order of three is applied and due to the weak excitation of the temperature, higher order regressors with the intake temperature are manually cancelled. Thus, the set of potential regressors has  $n = 26$  elements.

### Kernel model

Since the applied Kernel model consists of the entire training dataset, one kernel function needs to be evaluated for each training point. The applied kernel function is a Gaussian, as is also applied for LOLIMOT. Hence, the required FLOPs to evaluate the kernel functions in eq. (6.9) are identical to eq. (6.13), with  $M_L$  being the number of data points. For a local model there are in average  $M_L = 218$  data points in the training set. Besides the calculation of the kernel function, there are  $M_L$  multiplications with the measured output value and  $M_L - 1$  additions to calculate eq. (6.9). Therefore, an evaluation of  $m_{\text{nox},j}(\mathbf{u})$  with the Kernel model requires the following FLOPs,

$$\begin{aligned} + & : (2p + 1) \cdot M_L \\ * & : (2p + 2) \cdot M_L + 1 \\ \mathbf{exp} & : M_L \end{aligned} \quad (6.15)$$

with  $M_L = 218$  and the number of local model inputs  $p = 4$ .

### Comparison

For the global-local model structure, as in eq. (6.1), at most four of the 21 local models are valid and need therefore be evaluated for a global model simulation. The number of FLOPs required for one evaluation of  $\dot{m}_{\text{nox}}(\mathbf{z}, \mathbf{u})$  is therefore the number of FLOPs required to evaluate  $\dot{m}_{\text{nox},j}(\mathbf{u})$  multiplied by four plus the evaluation of one look-up table, eq. (6.11), over the engine operation point ( $\mathbf{z}^T = n_{\text{eng}}, u_{\text{inj}}$ ),

$$\text{FLOPs}(\dot{m}_{\text{nox}}(\mathbf{z}, \mathbf{u})) = 4 \cdot \text{FLOPs}(\dot{m}_{\text{nox},j}(\mathbf{u})) + \text{FLOPs}(\Phi(\mathbf{z})). \quad (6.16)$$

The latter is necessary to determine the weights for the local models. The resulting FLOPs are summarised in Table 6.2 for the four model structures. It shall be noted that an ECU with floating-point arithmetic is assumed for these results. The results for an ECU with fixed-point arithmetic are however assumed to be similar.

The Kernel model is only limitedly suited for an implementation in common ECUs, as it requires the highest computing time. Its computing time directly depends on the number of training values, but it might be reduced by choosing alternative Kernel functions or alternative Kernel methods like support vector machines [179]. The look-up tables on the other hand require the lowest number of FLOPs. LOLIMOT needs about three times of the look-up tables computing time, but requires additional evaluations of the exponential function. The adaptive polynomial approach LOPOMOT

**Table 6.2:** Required FLOPs for one simulation of the global-local  $\text{NO}_x$  model,  $\dot{m}_{\text{nox}}(\mathbf{z}, \mathbf{u})$ . It is distinguished between multiplications '\*', additions '+' and evaluations of an exponential function 'exp'. The four regarded model structures are each applied for the local models  $\dot{m}_{\text{nox},j}(\mathbf{u})$ , see eq. (6.1).

	+ operator	* operator	exp evaluation	sum of operations
look-up tables	143	164	-	307
LOLIMOT	455	485	34	974
LOPOMOT	115	216	-	331
Kernel model	6991	7864	872	15727

requires only slightly more FLOPs than the look-up table structure and is therefore also suited for an implementation in common ECUs.

The number of FLOPs is important for an implementation in the ECU. For other applications, also the time required for model training is of interest. This is the highest for the Kernel model, since a non-linear optimisation algorithm is applied to determine the bandwidth parameters. The fastest training is given for the look-up table approach, if the search of a suited look-up table structure and the determination of a regularisation parameter is not considered. This can increase the training time significantly such that the LOLIMOT structure might be faster to identify. Its tree construction algorithm is a fast heuristic search for suited model partitions and the determination of the model parameters is given by a least squares algorithm. LOPOMOT requires in general less partitions than LOLIMOT, but needs additional computing time for the selection of regressors. This additional time depends on the number of potential regressors and the applied selection algorithm. The model training is generally more extensive than for LOLIMOT.

### 6.1.3 Memory demand

The required memory to store the models is in the following regarded by the number of model parameters. Besides the parameters identified at model training, there are additional parameters necessary for the model structures.

For the look-up tables, additional parameters are required to define the grid point positions. If an equidistant grid point distribution is applied, this reduces to the discretisation of each input. For the regarded example, the distributions are identical for each local look-up table, why the additional parameters are counted for only one local model. For LOLIMOT, the standard deviations  $\sigma_j^2$  and the centres  $\mathbf{z}_{0,j}$  of each model partition are stored. Since these are vectors, there are  $p$  parameters stored for each standard deviation and each model centre. For LOPOMOT, the maximum and minimum values of the training data are stored for each local model. These are required to limit the input and output values of the polynomials to avoid unreasonable extrapolation results, see

Sect. 3.3. The Kernel model requires the entire training dataset. These are  $p$  input values and one output value for each of the 4584 measurements.

The numbers of stored parameters for the applied model structures are summarised in Tab. 6.3. The look-up table structure possesses the highest number of parameters to be identified. Due to the quadratic input space of look-up tables and the equal distribution of grid points in all local models, a large part of the parameters are however not in the drivability space. To identify reasonable values for these parameters, the regularisation as introduced in sect. 2.1.2 is applied. The distribution of grid points can also be adapted to the local models. This would however increase the number of additional parameters, since then each distribution need to be stored. For the applied equal distribution, the number of additional parameters can be neglected.

**Table 6.3:** Required memory for the different model structures to model  $\dot{m}_{\text{nox}}(\mathbf{z}, \mathbf{u})$ . Besides the identified parameters, additional parameters need to be stored for the model structures.

	identified parameters	additional parameters	sum of parameters
look-up tables	6300	40	6340
LOLIMOT	880	1408	2288
LOPOMOT	546	210	756
Kernel model	84	22920	23004

The LOLIMOT structure requires significantly less parameters for identification. This is since the models are identified on the reachable input space and an adaption is performed for each local model by the tree construction algorithm. Thus, considerably less redundant regressors are necessary than for the look-up table structure. This is however at the expense of additional parameters for the model structure and a more complicated interpretability of parameters.

LOPOMOT adapts to the non-linearity of the process by a selection of significant regressors. The local models are polynomials and are therefore more complex than the local models of LOLIMOT. Thus, no further partitions are necessary, which reduces the number of additional parameters. The additional parameters for LOPOMOT are required to prevent the local models from tending to infinity in the extrapolation regions.

For the Kernel model, only the bandwidths of the Kernel functions have to be identified. The bandwidths are individual identified for each local model in each of the four dimensions. Besides these small number of identified parameters, the Kernel model possesses the highest number of additional parameters, consisting of the entire training dataset with one coordinate for each dimension. Thus, the Kernel model has in total the highest memory demand compared to the other model structures.

#### 6.1.4 Conclusions

All model structures are able to model the globally modelled NO<sub>x</sub> emissions on the training data, but the Kernel model and the look-up table structure perform a bit worse on the validation data.

Regarding the look-up table structure, this is because of the high dimensional relation of the  $\text{NO}_x$  emissions. Look-up tables are mostly limited to two dimensions, why the accuracy of a look-up table structure decreases with an increasing dimension. On the other hand, the look-up table structure is the fastest to compute, but the advantages over LOPOMOT are relatively small. LOLIMOT requires about three times of the computing time, but suffers from additional evaluations of the exponential function. For an implementation in the ECU, the exponential function can be approximated by a discretisation or alternative weighting functions, such as polynomials [119], can be applied. The Kernel model possesses the highest evaluation time and can only limitedly be utilised for such applications in ECUs.

The ability to adapt to the process is reflected in the required memory, where LOLIMOT and LOPOMOT are the best performing structures. LOLIMOT adapts to the process by applying partitions of the input space and LOPOMOT by the selection of significant regressors for the local polynomial models. LOPOMOT has some advantages compared to LOLIMOT regarding the storing of additional parameters, since there are no partitions to be stored in this case. On the other hand, the look-up table structure can not adapt to the process. The discretisation of the look-up tables is the same for all local models, why a relative high number of parameters are required. An adaption of the discretisation would reduce the number of parameters to be identified and might also increase the model accuracy. However, the number of additional parameters and the computing time would increase at the same time.

All in all, for the regarded  $\text{NO}_x$  model, the LOPOMOT structure shows the best performance. If the weighting function of the LOLIMOT structure is modified, this possesses similar qualities. The look-up table structure suffers from its fixed discretisation, but has the best interpretability of parameters. The Kernel model performs worse in all categories. It can model a process without any a-priori knowledge if enough measurements are at hand, but suffers from its non-linear identification and the laborious evaluation using the entire training dataset. The LOPOMOT structure and also the LOLIMOT structure are therefore suited for an implementation in modern ECUs. They are especially advantageous, if a high dimensional relation has to be mapped. They possess further benefits if information about the process, such as derivatives, are required. Look-up tables are on the other hand advantageous if a fine tuning of model parameters is of interest. These different properties are again summarised in Tab. 6.4.

Examples for control functions modelled by look-up tables are presented in the following for a torque model and an open loop control of the smoke limitation. For the torque model, the look-up table structure is applied because of the required inversion. For the smoke limitation, it is applied because of the good interpretability of look-up table parameters.

**Table 6.4:** Summarised properties for the model structures discussed for the example of a  $\text{NO}_x$  emission model. The model structures are compared with regard to the categories model accuracy, computing time, memory demand and interpretability/ invertibility.

	model accuracy	computing time	memory demand	interpretability/ invertibility
look-up tables	o	++	-	+
LOLIMOT	++	+	+	-
LOPOMOT	++	++	++	-
Kernel model	++	-	-	-

## 6.2 Implementation of a torque model in the ECU

Several control functions in the ECU require a torque model,

$$\hat{M}_{\text{eng}} = f(n_{\text{eng}}, u_{\text{inj}}, \dots), \quad (6.17)$$

while the inverse relation from the engine torque to the required injection quantity is of interest for some other applications

$$u_{\text{inj}} = f^{-1}(\hat{M}_{\text{eng}})|_{n_{\text{eng}}, \dots}. \quad (6.18)$$

The prediction of the engine torque is needed because of the torque oriented control structure of modern ECUs [38]. In a torque oriented control structure, several requests, such as the drivers pedal input or the energy demands of auxiliaries, are formulated as torque requests. These requests are added and restricted by limitations. For such a structure, a torque model is for example needed to translate a fuel limitation, such as the one given by the smoke limitation, into a torque limitation. Once the torques are added and limited, the ECU needs to adjust the required torque by the corresponding fuel quantity. Therefore, the inverse relation from the torque to the injection quantity is of interest, for what an inverse torque model is required.

This inverse relation can either be calculated by a separate model  $g(\cdot)$ , determining the injection quantity in dependence on the required engine torque and other model influences,

$$u_{\text{inj}} = g(\hat{M}_{\text{eng}}, n_{\text{eng}}, \dots), \quad (6.19)$$

or by an inversion of the torque model as in eq. (6.18). A model for the inverse relation as in eq. (6.19) can be derived from a separate identification, applying the torque as model input and the injection quantity as model output. It has the advantage that the injection quantity is directly determined by an evaluation of  $g(\cdot)$ , but has the drawback that deviations between the predicted torque and its inverse calculated injection quantity can occur, since the inverse model is only an approximation  $g \approx f^{-1}$ . These deviations are even more problematic, if the torque model  $f$  is adapted during engine operation. Then the inverse model  $g$  needs to be adapted similarly, which

is a complicated task. Since already small deviations between the predicted torque and the corresponding injection quantity have negative effects on many control functions, the inversion of the torque model, as in eq. (6.18), is recommended for an application in the ECU.

To calculate the inverse of  $f$ , an injective mapping from  $u_{\text{inj}}$  to  $\hat{M}_{\text{eng}}$  is required such that a unique injection quantity corresponds to a requested torque. This constraint needs already be ensured at identification of the torque model.

Due to this injective property and the required inverse of the torque model, the possible model structures are limited. The local polynomial model, as presented for the engine torque in Chap. 3, can be inverted locally but it is complicated to guarantee the injective mapping for the global-local polynomial structure. Also the LOLIMOT and the Kernel model are complicated to be inverted. Hence, if an inversion is required, look-up tables are applied. In the following, a global look-up table structure is presented to model the engine torque. To include the major influences on the engine torque, a combination of several two-dimensional look-up tables, as introduced in Sect. 2.1.3, is applied. The presented structure consists of a base look-up table, multiplicative correction look-up tables and an additive influence. Given this model structure, the injective mapping is investigated. Finally, the inversion of the model torque is presented.

### 6.2.1 Look-up table structure for the torque model

The utilised look-up table structure is based on the physical influences on the engine torque. It is therefore especially not determined by trial and error, as applied for the NO<sub>x</sub> model in the previous section. The utilised model structure is depicted in Fig. 6.2.

The engine torque primarily depends on the engine operation point  $(n_{\text{eng}}, u_{\text{inj}})$ . For these influences a base look-up table is identified, which models the inner torque  $\hat{M}_{\text{in}}$ . It is assumed that all other influences are in an optimal state. Deviations from these optimal states are modelled as efficiency factors  $\eta_i$ , [38]. They are also modelled by two-dimensional look-up tables depending on the engine speed and one other influence. The major influences are  $\varphi_{Q50}$ ,  $p_{2i}$ ,  $m_{\text{air}}$  and  $T_{2i}$ . Depending on how accurate the torque model need to be, more or less correction look-up tables are applied. After multiplying  $\hat{M}_{\text{in}}$  by the efficiency factors  $\eta_i$ , the effective inner torque  $\hat{M}_{\text{in,eff}}$  results. From this the friction torque  $\hat{M}_{\text{fric}}$  is subtracted such that the torque on the crankshaft is given by

$$\hat{M}_{\text{eng}}(n_{\text{eng}}, u_{\text{inj}}, m_{\text{air}}, p_{2i}, \varphi_{mi}, T_{2i}) = \hat{M}_{\text{in,eff}} - \hat{M}_{\text{fric}} \quad (6.20a)$$

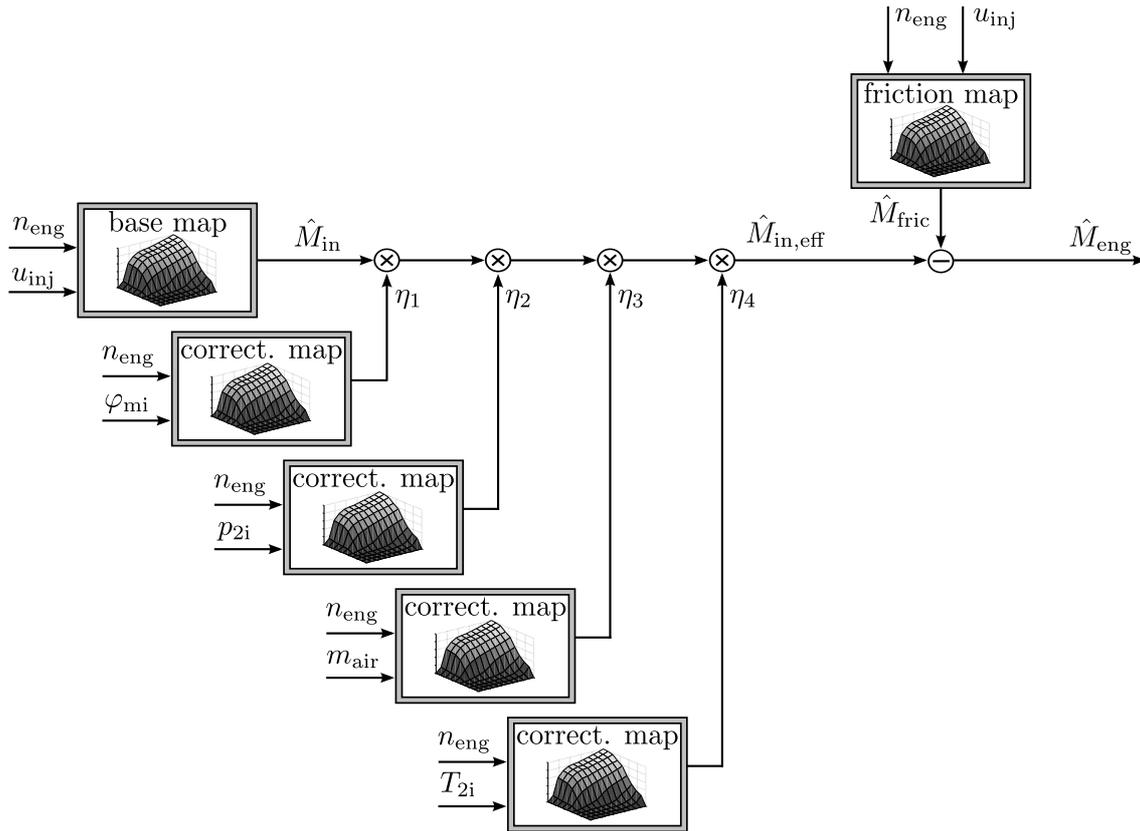
$$= \hat{M}_{\text{in}} \eta_1 \eta_2 \eta_3 \eta_4 - \hat{M}_{\text{fric}}. \quad (6.20b)$$

with

$$\hat{M}_{\text{in}} = f_{\hat{M}_{\text{in}}}(n_{\text{eng}}, u_{\text{inj}}), \quad \hat{M}_{\text{fric}} = f_{\hat{M}_{\text{fric}}}(n_{\text{eng}}, u_{\text{inj}}), \quad (6.21a)$$

$$\eta_1 = f_{\eta_1}(n_{\text{eng}}, \varphi_{mi}), \quad \eta_2 = f_{\eta_2}(n_{\text{eng}}, p_{2i}), \quad (6.21b)$$

$$\eta_3 = f_{\eta_3}(n_{\text{eng}}, m_{\text{air}}), \quad \eta_4 = f_{\eta_4}(n_{\text{eng}}, T_{2i}). \quad (6.21c)$$



**Figure 6.2:** Multiplicative and additive look-up table structure for a mean value model of the engine torque with six input variables. The modelled inner torque  $\hat{M}_{in}$  is multiplied by the correction factors  $\eta_1$  to  $\eta_4$  to the effective inner torque  $\hat{M}_{in,eff}$ . Then the friction torque  $\hat{M}_{fric}$  is subtracted, to finally get the modelled torque at the crankshaft  $\hat{M}_{eng}$ .

The friction torque  $\hat{M}_{fric}$  effects the inner torque additively and is also modelled as a function of the engine operation point  $(n_{eng}, u_{inj})$ . It describes the differences between the measured torque at the crankshaft and the indicated torque calculated with in-cylinder pressure sensors and comprises the friction of moving components. The influence of the engine temperature on the friction torque is disregarded, but can easily be added to the structure by another look-up table. Since the friction torque is modelled as a function of the injection quantity, this model needs also be regarded for model inversion.

The discretisations of the look-up tables are chosen a-priori such that the determination of the model parameters is the only task for model identification. Two approaches are presented in the following to identify the model parameters. The first approach identifies each look-up table *separately*, whereas the second approach identifies all look-up tables with one dataset *simultaneously*. In both cases the look-up table of the friction torque is identified in a previous identification, independent from the identification of the other look-up tables.

- a) For the *separate identification*, at first the optimal states  $\eta_i = \eta_{opt} = 1$  are determined in each engine operation points. The corresponding torques are then stored in the base look-

up table  $\hat{M}_{\text{in}}(n_{\text{eng}}, u_{\text{inj}})$ . Measurements for the torque can either be taken from the measured torque at the crankshaft corrected by the modelled friction torque, or from the indicated torque calculated from in-cylinder pressure sensors. Thereafter, the efficiency factors, e. g.  $\eta_1 = f(n_{\text{eng}}, \varphi_{Q50})$ , are successively calibrated by measurements on their grid points.

b) For a *simultaneous identification* of the look-up tables, all influences on the torque,

$$\mathbf{u}^T = [n_{\text{eng}}, u_{\text{inj}}, m_{\text{air}}, p_{2i}, \varphi_{\text{mi}}, T_{2i}] \quad (6.22)$$

are excited within one dataset. Again, either the friction torque can be added to the measured torque at the crankshaft or the indicated torque is applied. Thus, measurements are obtained for the effective torque  $M_{\text{in,eff}} = f(n_{\text{eng}}, u_{\text{inj}}, m_{\text{air}}, p_{2i}, \varphi_{\text{mi}}, T_{2i})$  in dependence of all model inputs, see eq. (6.22). With this dataset, at first a base look-up table  $\hat{M}_{\text{in}}^*(n_{\text{eng}}, u_{\text{inj}})$  is identified, for which  $\hat{M}_{\text{in}}^*$  is approximated by a regression with regard to  $(n_{\text{eng}}, u_{\text{inj}})$ . From the multiplicative residuals, see eq. (2.21), the correction look-up tables  $\eta_1^*$  to  $\eta_4^*$  are trained successively, again by regressions. Due to these regressions, the correction look-up tables possess values greater and smaller than one. To interpret these as efficiency factors, they are normed to one after identification,

$$\eta_1 = \frac{\eta_1^*}{\max \eta_1^*}, \quad \eta_2 = \frac{\eta_2^*}{\max \eta_2^*}, \quad \eta_3 = \frac{\eta_3^*}{\max \eta_3^*}, \quad \eta_4 = \frac{\eta_4^*}{\max \eta_4^*} \quad (6.23)$$

and the factor is assigned to the base look-up table

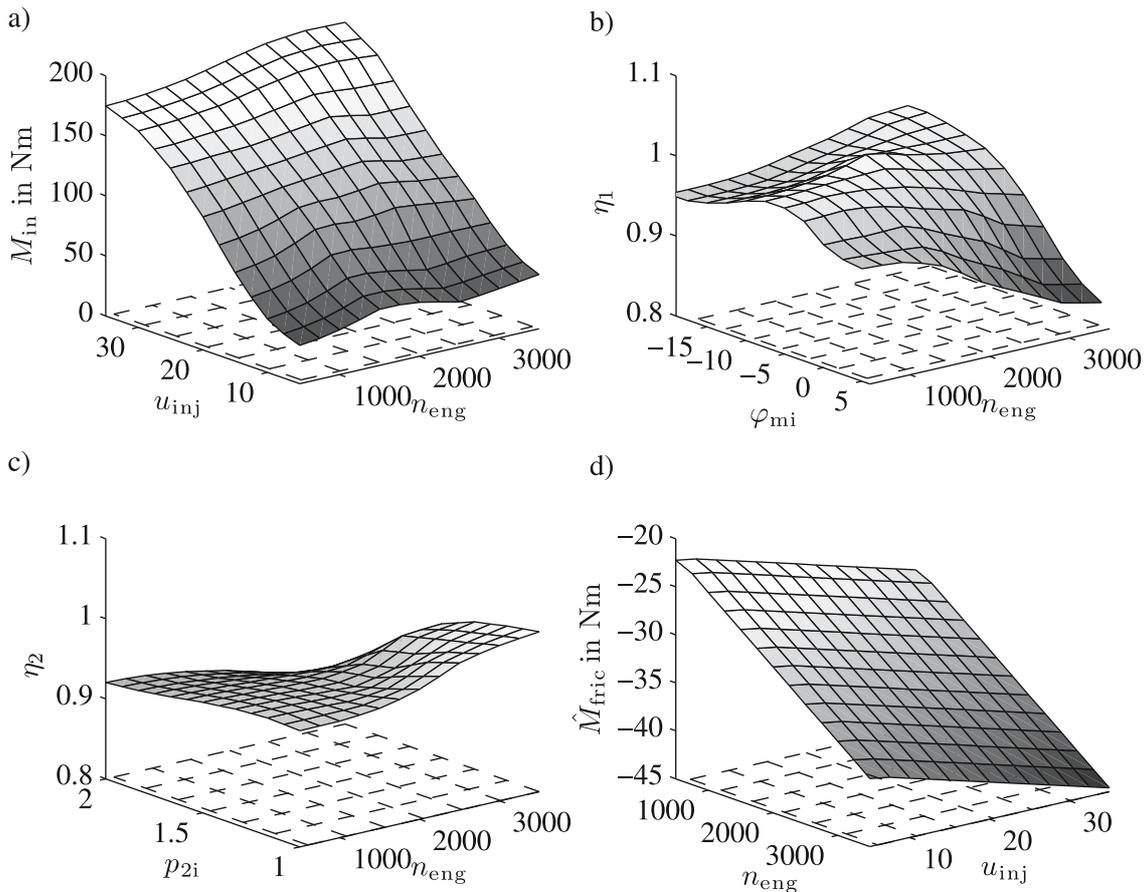
$$\hat{M}_{\text{in}}(n_{\text{eng}}, u_{\text{inj}}) = \hat{M}_{\text{in}}^*(n_{\text{eng}}, u_{\text{inj}}) \cdot (\max \eta_1^*) \cdot (\max \eta_2^*) \cdot (\max \eta_3^*) \cdot (\max \eta_4^*). \quad (6.24)$$

With either approach, the torque model as in eq. (6.20) results. Applying approach b) for identification with the dataset as presented in Sect. 3.3, the quality of the torque model is  $R^2 = 0.989$  and  $R^2 = 0.972$  on training respectively validation data. It is therefore in the range of the LOPO-MOT model structure as presented in Chap. 3. The identified base look-up table is plotted together with the first correction look-up table in Fig. 6.3. For each look-up table, a linear regularisation as presented in Sect. 2.1.2 is applied.

### 6.2.2 Injective mapping constraint for the torque model

The presented model simulates the engine torque. However, some control functions in the ECU require the inverse relation to the injection quantity  $u_{\text{inj}}$ . To invert the look-up table with regard to  $u_{\text{inj}}$ , an injective mapping from  $u_{\text{inj}}$  to  $M_{\text{eng}}$  is required. An injective mapping implies that the model is strictly monotone increasing or strictly monotone decreasing in  $u_{\text{inj}}$ . This constraint can be formulated for a two-dimensional look-up table with regard to the look-up table heights  $w_{i,j}$ . The look-up table is strictly monotone increasing in its second coordinate  $u_{\text{inj}}$ , if the following holds for all grid points  $i, j$  in the look-up table

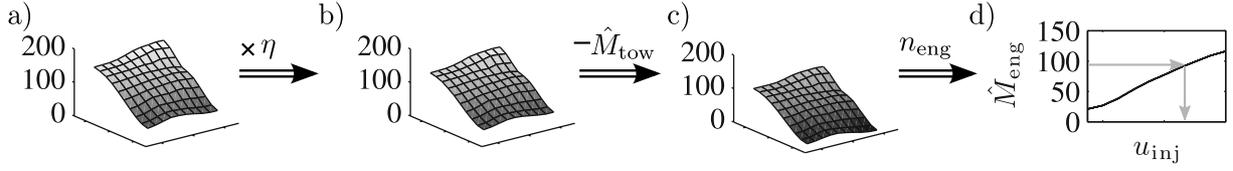
$$w_{i,j} < w_{i,j+1}. \quad (6.25)$$



**Figure 6.3:** a) Identified base look-up table for the inner torque  $\hat{M}_{in}(n_{eng}, u_{inj})$ , b) first correction look-up table for the crank angle of main injection  $\eta_1(n_{eng}, \varphi_{mi})$ , c) second correction look-up table for the intake pressure  $\eta_2(n_{eng}, p_{2i})$  and d) the look-up table for the friction torque  $\hat{M}_{fric}(n_{eng}, u_{inj})$ . Correction look-up tables can be interpreted as efficiency factors. The base and the correction look-up tables are smoothed by a regularisation. The friction torque is modelled as a linear function on its model inputs.

A strictly monotone decreasing property can be formulated analogue. This constraint needs to be checked for a given look-up table. If it is not satisfied, the model parameters need to be adjusted with regard to this constraint. For the regarded torque model, several look-up tables are combined, see Fig. 6.2. Therefore, the injective mapping has to be checked for the entire model structure.

To investigate the injective mapping from  $u_{inj}$  to  $\hat{M}_{eng}$ , the look-up tables which depend on the injection quantity  $u_{inj}$  are regarded. These are the base look-up table  $\hat{M}_{in}$  and the look-up table for the friction torque  $\hat{M}_{fric}$ . The base look-up table is strictly monotone increasing in  $u_{inj}$ , as can be seen in Fig. 6.3 a). This monotony also holds for the effective engine torque  $\hat{M}_{in,eff}$ , which is derived from a multiplication of  $\hat{M}_{in}$  by the efficiency factors  $\eta_i$ . The efficiency factors are regarded as constants for an inversion with regard to  $u_{inj}$ . Then the friction torque is subtracted. Since the model of the friction torque is strictly monotone decreasing in  $u_{inj}$ , a subtraction from the strictly



**Figure 6.4:** Inversion of the look-up table structure, starting with a) the base look-up table  $\hat{M}_{in}(n_{eng}, u_{inj})$ . This is multiplied by the efficiency factor, here  $\eta = 0.89$ , to b). After subtraction of the friction torque to c), the resulting look-up table is simplified to a characteristic curve d) with regard to the current engine speed  $n_{eng}$ . The required injection quantity  $u_{inj}$  for a given torque request  $M_{eng}$  is then determined by an inverse evaluation of the characteristic curve.

monotone increasing model  $\hat{M}_{in,eff}$  results again in a strictly monotone increasing model. Hence,

$$\hat{M}_{eng}(n_{eng}, u_{inj,1}, m_{air}, p_{2i}, \varphi_{mi}, T_{2i}) < \hat{M}_{eng}(n_{eng}, u_{inj,2}, m_{air}, p_{2i}, \varphi_{mi}, T_{2i}) \quad (6.26)$$

holds for  $u_{inj,1} < u_{inj,2}$ . Thus, if  $n_{eng}, m_{air}, p_{2i}, \varphi_{mi}$  and  $T_{2i}$  are given, an unique injection quantity can be determined for an engine torque request.

### 6.2.3 Inversion of the torque model

To determine the injection quantity for a requested torque, the look-up table structure from Fig. 6.2 needs to be inverted. At first, the efficiency factors, which are regarded as constants, are calculated and summarised to one efficiency factor

$$\eta|_{n_{eng}, m_{air}, p_{2i}, \varphi_{mi}, T_{2i}} = \eta_1|_{n_{eng}, \varphi_{mi}} \eta_2|_{n_{eng}, p_{2i}} \eta_3|_{n_{eng}, m_{air}} \eta_4|_{n_{eng}, T_{2i}}. \quad (6.27)$$

The efficiency factors are regarded as constants, since they do not depend on the injection quantity  $u_{inj}$ . The base look-up table  $\hat{M}_{in}(n_{eng}, u_{inj})$  can then be multiplied by the efficiency factor  $\eta$  to get the effective engine torque look-up table  $\hat{M}_{in,eff}(n_{eng}, u_{inj})$ . From this, the look-up table of the friction torque  $\hat{M}_{fric}(n_{eng}, u_{inj})$  is subtracted. Thus, only one look-up table results, which depends on the engine speed and the injection quantity

$$\hat{M}_{eng}|_{m_{air}, p_{2i}, \varphi_{mi}, T_{2i}}(n_{eng}, u_{inj}) = \eta|_{n_{eng}, m_{air}, p_{2i}, \varphi_{mi}, T_{2i}} \hat{M}_{in}(n_{eng}, u_{inj}) - \hat{M}_{fric}(n_{eng}, u_{inj}). \quad (6.28)$$

The look-up table  $\hat{M}_{eng}(n_{eng}, u_{inj})$  can further be reduced to a characteristic curve with regard to the current engine speed such that the model structure of eq. (6.20) simplifies to

$$\hat{M}_{eng}|_{n_{eng}, m_{air}, p_{2i}, \varphi_{mi}, T_{2i}}(u_{inj}) = \eta|_{n_{eng}, m_{air}, p_{2i}, \varphi_{mi}, T_{2i}} \hat{M}_{in}|_{n_{eng}}(u_{inj}) - \hat{M}_{fric}|_{n_{eng}}(u_{inj}). \quad (6.29)$$

The characteristic curve  $\hat{M}_{eng}(u_{inj})$  is then strictly monotone increasing in  $u_{inj}$ . Hence, the required injection quantity can be determined from a given torque request by an inverse evaluation of the characteristic curve. The discussed individual steps to invert the look-up table structure are also depicted in Fig. 6.4.

The inversion can be accelerated, if the base look-up table and the look-up table of the friction torque are reduced to characteristic curves, before they are subtracted from each other. The computing time can further be reduced, if only the relevant regions of the look-up tables are regarded. However, these details concern a final implementation in the ECU and are not further discussed here.

### 6.3 Open loop control of the smoke limitation

The model-based smoke limitation, presented in Sect. 5.4, has the advantage that the main influences on the formation of soot are covered by the applied soot model. Thus, no additional calibration effort is necessary. A drawback is that an interpretation of the model parameters is difficult, why a fine tuning of this model-based smoke limitation is a complicated task. Furthermore, a fault-free functionality is laborious to check and the numerical inversion, as presented in Sect. 5.4.2, requires several computing steps.

To maintain the positive properties and avoid the negative, an open loop control of the smoke limitation is derived from the model-based approach in the following. Therefore, the inverse of the soot model is approximated by a separate model. Small deviations between the soot model and the approximated inverse are unproblematic, since these deviations have no effects on other control functions. Since look-up tables are best suited for calibration tasks, a look-up table structure, which consists of several two-dimensional look-up tables, is applied.

The model-based smoke limitation in Sect. 5.4 determines the maximum injection quantity  $u_{\text{inj,limit}}$  with regard to the predefined soot limit  $c_{\text{soot,limit}} = 15 \text{ mg/m}^3$ , see eq. (5.7). With the engine states being regarded as constants, see eq. (5.9), the condition for the inversion is given by

$$c_{\text{soot}}(u_{\text{inj,limit}}) \Big|_{n_{\text{eng}}, x_{2i}, m_{\text{eng,in}}, p_{2i}, \varphi_{\text{mi}}, T_{2i}} \stackrel{!}{=} c_{\text{soot,limit}}. \quad (6.30)$$

The objective is to approximate the inverse by a function  $\tilde{g} \approx c_{\text{soot}}^{-1}$  such that

$$u_{\text{inj,limit}} = \tilde{g} \left( n_{\text{eng}}, x_{2i}, m_{\text{eng,in}}, p_{2i}, \varphi_{\text{mi}}, T_{2i} \right) \Big|_{c_{\text{soot,limit}}}. \quad (6.31)$$

This is in general possible, but for a better comparison with common smoke limitations, the indirect limitation of the injection quantity via a limit for the air to fuel ratio  $\lambda$  is applied. Hence, the following relation needs to be approximated

$$\lambda_{\text{limit}} = g \left( n_{\text{eng}}, x_{2i}, m_{\text{eng,in}}, p_{2i}, \varphi_{\text{mi}}, T_{2i} \right) \Big|_{c_{\text{soot,limit}}}. \quad (6.32)$$

The limit of the injection quantity follows then by eq. (A.5),

$$u_{\text{inj}} = \frac{x_{2i} m_{\text{eng,in}}}{L_{\text{st}} \rho_{\text{Diesel}} \lambda}. \quad (6.33)$$

The applied look-up table structure to approximate  $g$  is derived from trial and error. Therefore, a dataset is created, which excites the inverse model  $g$  in all its inputs

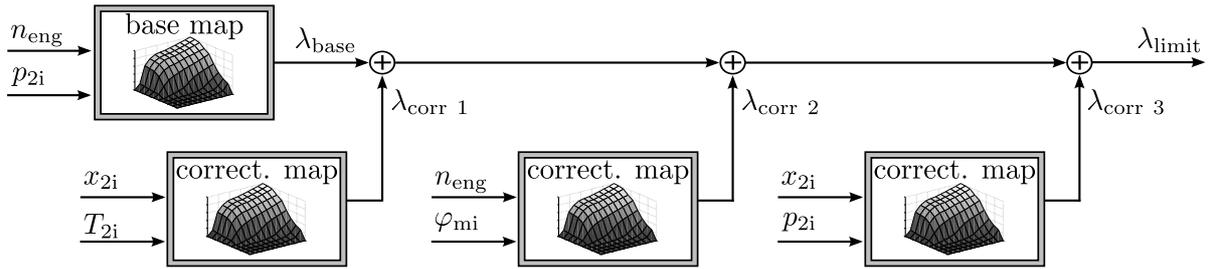
$$\mathbf{u}^T = [n_{\text{eng}}, x_{2i}, m_{\text{eng,in}}, p_{2i}, \varphi_{\text{mi}}, T_{2i}]. \quad (6.34)$$

Since the dataset presented in Sect. 3.3 excites these inputs, these measurements are utilised. For each of these measurements, the limit for the injection quantity  $u_{\text{inj,limit}}$  is determined by the model-based smoke limitation from Sect. 5.4. With this limit and the corresponding values for  $x_{2i}$  and  $m_{\text{eng,in}}$ , given by the regarded measurements, the limit for the air to fuel ratio is determined by

$$\lambda_{\text{limit}} = \frac{x_{2i} m_{\text{eng,in}}}{L_{\text{st}} \rho_{\text{Diesel}} u_{\text{inj,limit}}}, \quad (6.35)$$

see eq. (6.33). Thus, a dataset with a six dimensional input, see eq. (6.34), and a scalar output as calculated in eq. (6.35) results.

Given this dataset  $(\mathbf{u}(i), \lambda_{\text{limit}}(i))$ , at first the base look-up table is determined by a regression. Therefore, the two inputs of eq. (6.34) are selected that approximate the six dimensional relation in eq. (6.32) best. These are  $n_{\text{eng}}$  and  $p_{2i}$ . Given this base look-up table, the residuals are determined by eq. (2.20), since an additive structure is applied. Then the first correction look-up table is determined. Again, all feasible input combinations are tested and the inputs with the best approximation qualities are selected, which are  $x_{2i}$  and  $T_{2i}$ . With this correction look-up table, the residuals are calculated and two more correction look-up tables are identified analogously. Thus, the model structure as presented in Fig. 6.5 results.



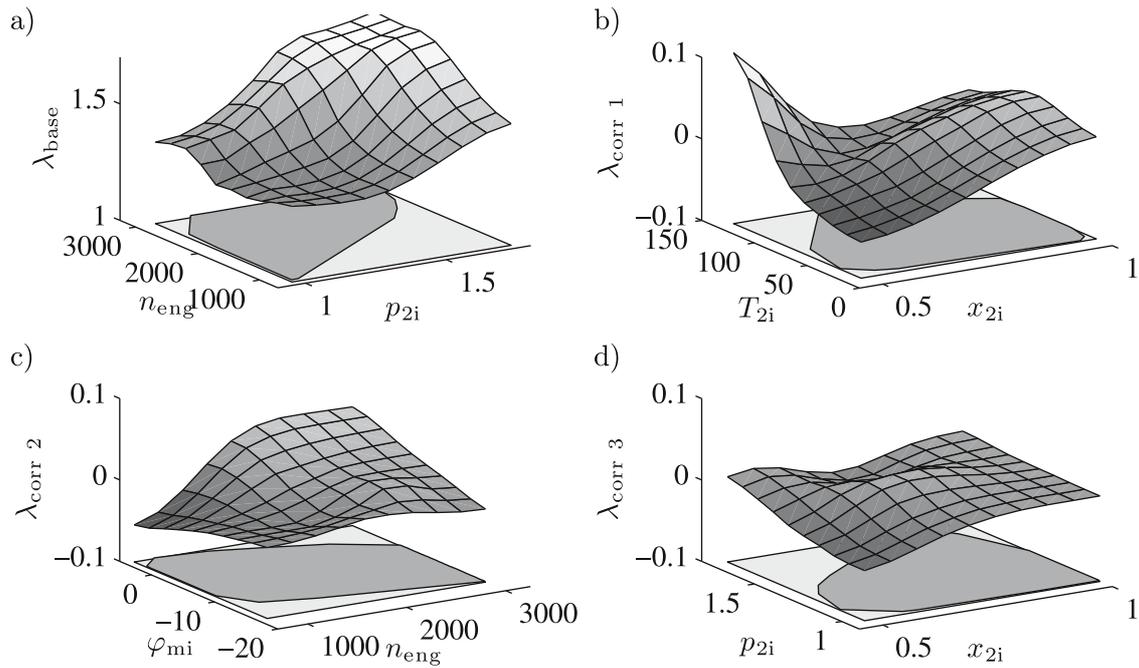
**Figure 6.5:** Look-up table structure for an open loop control of the smoke limitation. The injection quantity limit  $u_{\text{inj,limit}}$  is determined by the modelled air to fuel ratio limit  $\lambda_{\text{limit}}$  in connection with eq. (6.33). Inputs of the look-up tables are derived by trial and error.

The inputs of the look-up tables are discretised in ten equidistant steps. The maximum and minimum values are determined by the observed maximum and minimum values from training data. The input  $m_{\text{eng,in}}$  is not selected for any of those look-up tables. Hence, five inputs are applied to model  $g$ , but the gas mass in the engine  $m_{\text{eng,in}}$  affects the injection quantity limit by eq. (6.33).

Given this look-up table structure, the approximation of eq. (6.32) is given by

$$\lambda_{\text{limit}} = \lambda_{\text{base}}(n_{\text{eng}}, p_{2i}) + \lambda_{\text{corr 1}}(x_{2i}, T_{2i}) + \lambda_{\text{corr 2}}(n_{\text{eng}}, \varphi_{\text{mi}}) + \lambda_{\text{corr 3}}(x_{2i}, p_{2i}). \quad (6.36)$$

The identified look-up tables are presented in Fig. 6.6. The base look-up table gives the basis value for the air to fuel ratio limit, which is always above 1.0. The highest variabilities for the limit value are covered by this look-up table. The influences of the correction look-up tables reduces from one correction look-up table to the other such that the influence of  $\lambda_{\text{corr 3}}$  is relatively small. Hence, an



**Figure 6.6:** Identified look-up tables for the open loop control of the smoke limitation, see eq. (6.36). The base look-up table gives the basis value for the air to fuel ratio limit, which is corrected by the three subsequent correction look-up tables. The light grey areas indicate the input spaces of the look-up tables and the dark grey areas the regions where measurements are available.

approximation with further correction look-up tables would not influence the air to fuel ratio limit significantly.

The input spaces of the look-up tables are depicted by the light grey areas in the plots and the regions covered with measurements by the dark grey areas. For the base look-up table, the measurements cover only about half of the input space such that a major region of the look-up table is not reachable in practice. To be able to identify the parameters in these regions without measurements, the regularisation presented in Sect. 2.1.2 is applied. Depending on the distribution of grid points and the shape of the drivability space, there might be major parts of the look-up table without practical relevance. These parameters need then to be stored for an application of the look-up table, but they have no contribution to the model quality.

The main advantage of the look-up table structure is the intuitive interpretation of parameters. Regarding the first correction look-up table, there are unreasonable values in the extrapolation region, see  $\lambda_{\text{corr}1}$  for low  $x_{2i}$  and high  $T_{2i}$  values. These parameters can easily be adjusted manually such that the functionality of look-up tables can be ensured also for extrapolation regions. Furthermore, the derived look-up tables can be optimised by a fine tuning of individual look-up table heights.

## 6.4 Summary

Look-up tables, LOLIMOT, LOPOMOT and a Kernel model are compared to each other with regard to model accuracy, computing time and required memory using the example of a  $\text{NO}_x$  emission model. It shows that LOLIMOT and LOPOMOT are suited alternatives to look-up tables for an application in the ECU with floating-point arithmetic. They are superior to look-up tables regarding the model quality and also regarding the number of model parameters. The advantages are due to the adaption of the models to the non-linearities of the process. LOLIMOT adapts to the non-linearities by its axis orthogonal input space partitions and LOPOMOT by the local selection of polynomial regressors. Their computing time is slightly larger than for the look-up tables, whereas the differences are relatively small. They are therefore superior to look-up tables regarding the mapping of higher dimensional relations, whereas LOPOMOT performs slightly better than LOLIMOT.

Look-up tables are on the other hand well suited to map lower dimensional relations up to two inputs. Hence, look-up tables are advantageous for lower dimensional relations, whereas LOLIMOT and LOPOMOT require less parameters and are more accurate for higher dimensions. A Kernel model can only limitedly be applied on the ECU. Its computing time and model quality depends on the number and distribution of measurements. The performance of the Kernel model is for the regarded example the worst in all categories.

Regarding a torque model, it is mandatory to invert the model with regard to the injection quantity. The complex model structures are only limitedly suited for an analytic inversion, why a look-up table structure is recommended here. The look-up table structure is derived from physical knowledge applying efficiency factors and a model for the friction torque. This torque model shows comparable results to complex model structures. The influence of the injection quantity on the torque is investigated and a constraint for the invertibility is formulated such that the existence of an unique inverse is ensured. The calculation of this inverse is outlined. Because of the good interpretability of parameters, look-up tables are better suited than complex model structures if an inversion is required.

Another advantage of look-up tables, related to the good interpretability of parameters, is the possibility of a fine tuning. The model-based smoke limitation presented in Sect. 5.4 can be directly derived from the soot model. It possesses however the drawback that the model parameters are difficult to interpret and therefore complicated to adjust. To enable a fine tuning and ensure the functionality of such a control function, an open loop control based on look-up tables is presented for the smoke limitation. The look-up table structure and the calibration of the look-up tables follow directly from the model-based smoke limitation. The derived open loop control has the advantage of an easy calibration since the model parameters are interpretable.

## 7 Conclusions

The increasing complexity of modern combustion engines requires the utilisation of efficient calibration methods. A manual calibration on the test bed is laborious and it can not be guaranteed that an optimum is found. Because of the shorter development cycles and since test bed time is expensive, long test runs must be avoided. This is why model-based optimisation approaches became essential in recent years for the calibration of engine control. In this thesis, the model-based optimisation is presented from the identification of models to the implementation of control functions in the ECU. Several enhancements to the model-based optimisation process are presented.

To optimise the engine control, the combustion outputs  $\text{NO}_x$ , soot and the engine torque are modelled by a mean value model. Therefore, the measurement dynamics are separated from the process. These measurement dynamics are especially for the emission measurements significant because of the relative long travelling time through the exhaust pipe and connection tubes. With this separation of measurement dynamics, the process is further separated into a dynamical and a stationary part. The dynamics are introduced by the air path, more precisely by the inertia of the turbocharger, the travelling time in the egr system and the dynamics in the intake and outlet. The injection characteristics, such as the injection angles, can be adjusted cyclewise why no dynamics are considered here. The dynamics of the rail pressure are neglected. Applying the states of the air path at intake valve closing and the injection characteristics as model inputs, a stationary model structure is applied for the mean value model of the combustion process.

Depending on the regarded model output and also on the purpose of the model, various inputs are possible. These inputs are often redundant from an analytical point of view, but they differ in their non-linearity with regard to the applied output and therefore in their attained model quality. For the soot model it is superior to apply the gas composition after combustion as model input, whereas the air mass per cycle is more meaningful for the  $\text{NO}_x$  emissions. Further utilised inputs are the engine speed, the injection quantity, the intake pressure, the intake temperature and the crank angle of 50 % mass fraction burnt or the crank angle of main injection if no in-cylinder pressure sensors are available. Furthermore, a logarithmic transformation of the soot measurement is presented, which transforms the multiplicative measurement error into an additive error. The applied model structure is a global-local model structure, which combines several local models for fixed engine operation point to a global model by means of a bilinear interpolation.

For visualisation of these high-dimensional models, an intersection plot tool is introduced that shows the modelled output in dependence to one or two inputs. An intersection plot is applied to illustrate the  $\text{NO}_x$ -soot trade-off, which is a challenge for the optimisation of a Diesel engine. To analyse the dynamic quality of the combustion models, models for the measurement dynamics are necessary. Their dynamics highly depend on the volume flow rate of the exhaust such that variable dead times between 2.2 s and 3.2 s occur for the soot measurement for the employed setting at the test bed of the Institute of Automatic Control and Mechatronics. The simulation of the combus-

tion models together with the models of the measurement dynamics show then a good dynamic correlation for the measured and modelled emission signals.

The stationary optimisation assumes that all dynamic effects are decayed and is therefore performed with the combustion models alone. The stationary optimisation is divided into a local optimisation and a global optimisation. The objective function of the global optimisation is derived from the regarded emission standard. It minimises the bsfc over a test cycle with regard to constraints for the accumulated emissions over the test cycle. The weights for the local models are derived from the frequency of occurrence of their operation points in the test cycle. An objective function for the local optimisation can only hardly be derived from the given emission standard limits. Therefore, a weighted criterion of the emissions  $\text{NO}_x$  and soot is applied such that the trade-off between those emissions is observable for varying weights. Alternative criteria are stated. The results of the local optimisation serve as initial values for the global optimisation. Good initial values are required, since the global optimisation is a high dimensional non-linear constraint optimisation problem.

Influences of model uncertainties and series variations are regarded. These variations deteriorate the optimisation results such that an engine of a series production may violate the emission limits if the optimisation results are on a limit value. Uncertainties and series variations due to sensors and construction tolerances are therefore considered in the boundary condition such that the obtained results keep a safety distance to the limit values. The simulated series variations show no violation of the emission standards. This optimisation is referred to as robust global optimisation. It also contains a term for the additional dynamic emissions, which are in general not regarded by a stationary optimisation. Influences of the dynamic emissions are investigated by an overall engine simulation from the actuators to the combustion outputs. The stationary combustion model is therefore connected to a dynamic air path model. The additional dynamic emissions are relatively small compared to the influences of model uncertainties and series variations. The dynamic emissions further depend on the applied control structure.

Various engine closed loop controls are regarded. An overview of common air path controls is given and alternative designs, controlling the emissions directly, are discussed. The optimisation of an acceleration event shows the necessity of these controls. The regarded acceleration event is a step in injection quantity followed by a ramp in engine speed. For this acceleration event, the trajectories of the air path actuators (guide vanes of the turbocharger and valve of the egr system) are optimised. The regarded objective function penalises the deviation of the dynamic emissions from their stationary calibrated values. Except for a soot peak shortly after the step in injection quantity, the optimised dynamic emissions follow their desired values without a significant deviation. To avoid the soot peak at the beginning of the step, a model-based smoke limitation is presented. This limits the soot emissions but also the engine torque, why it should not be too stringent. The introduced model-based smoke limitation applies the presented soot model and performs a numerical inversion to determine the limit value for the injection quantity. Measurements from the engine test bed show the performance of this smoke limitation.

Finally, the implementation of control functions in the ECU is investigated. Four experimental model structures are compared to each other. These are a look-up table model, a LOLIMOT model, a LOPOMOT model and a Kernel model. Look-up tables are common model structures for an implementation in the ECU. These are mostly limited to two inputs, why possible combinations with several two-dimensional look-up tables are introduced to map higher dimensional relations. Furthermore, look-up tables are mostly limited to a lattice structure, why several grid points are only weakly or not excited at all. Therefore, a regularisation is introduced which penalises the curvature of a look-up table and enables the identification of not excited parameters.

LOLIMOT is a multi-model approach which models a process by the superposition of several linear or affine functions. LOPOMOT is an enhancement to LOLIMOT applying adaptive local polynomial models. The theoretical derivation of the selection of regressors is presented. Because of the more complex local models, LOPOMOT requires less model partitions than LOLIMOT for a comparable model quality. This is especially advantageous, if only few measurements are available. LOPOMOT adapts to the non-linearities of the process by a local selection of regressors and by the tree construction algorithm. The local selection of regressors is also advantageous if redundant regressors are in the model. A comparison to LOLIMOT shows the advantages of the further developments.

Furthermore, the Kernel model is introduced as an universal approximator. It models the process by a weighted summation of the observed outputs from training data and therefore depends on the size of the training data. The model structures are compared to each other with regard to model accuracy, computing time and required memory using the example of a global-local  $\text{NO}_x$  emission model. The Kernel model shows the worst performance. LOLIMOT and LOPOMOT are superior to look-up tables regarding the model accuracy and the required memory. Their computing time is about two to three times of the computing time of look-up tables. They can therefore be regarded as superior to look-up tables regarding the mapping of higher dimensional relations. Look-up tables have however the advantage that the parameters are well to interpret. These are therefore better suited if the regarded model needs to be inverted, as is shown using the example of a torque model. Look-up tables are also superior if the model parameters need to be adjusted by a test bed engineer. Therefore, a look-up table structure is derived for an open loop control of the smoke limitation from the presented model-based smoke limitation. Thus, the basic calibration of the smoke limitation is still derived utilising the soot model, but the fine tuning of model parameters is eased.

## Outlook

There is further potential for future research on the field of model-based optimisation, regarding the consideration of further model inputs, such as the splitting of injections or the rail pressure. Then, similar modelling procedures are applied, but the dimensionality increases, which also increases the measurement time. An online measurement strategy need then be applied to reduce the test bed time. Online strategies detect e. g. stationary conditions automatically such that the holding time for a stationary measurement can be reduced significantly. It is further possible to perform parts of

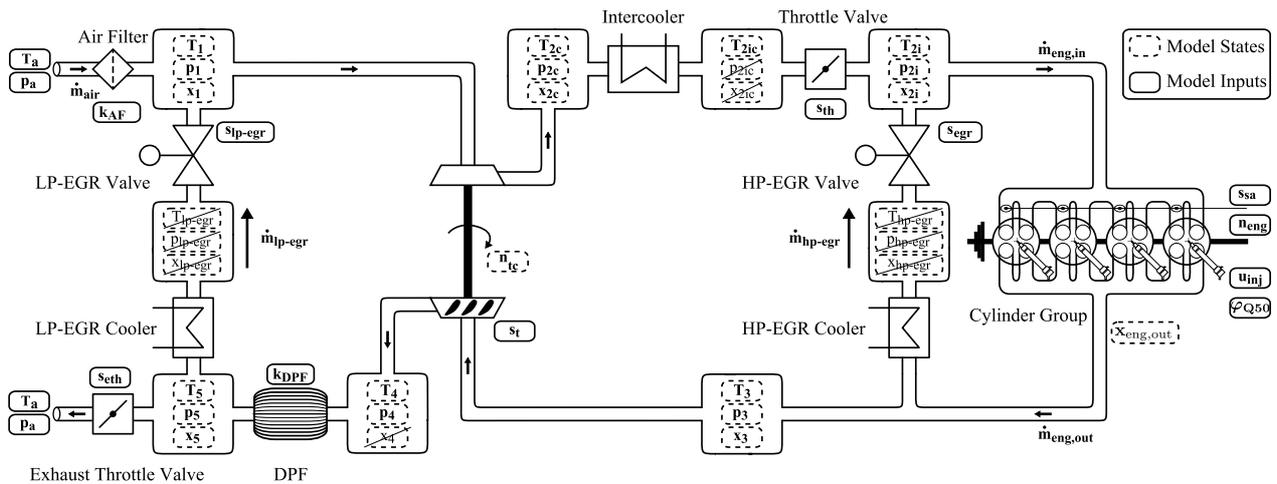
the modelling process parallel to the recording of measurements and adapt the measurement design to the model structure. Then a D-optimal design can be applied, which is not applied for this work to enable an unbiased comparison of model structures. Also parts of the optimisation process can be included into online strategies.

Besides this, the connection of the dynamic air path model to the stationary combustion model enables to investigate new advanced control strategies and compare them to common air path controls. Thus, new control design, such as the direct control of emissions, can be tested. Furthermore, dynamic optimisation results, as presented for the optimisation of the acceleration event, can be translated into feedforward controls.

Regarding the stationary optimisation process, the numerical optimisation algorithm can be adjusted to the optimisation problem. This accelerates the optimisation and is especially beneficial if several optimisation runs need to be performed. If the optimisation is accelerated, the smoothing of engine look-up tables can be integrated into the optimisation process.

## A Air Path Model

The air path is modelled with a lumped parameter approach and is kindly provided by Matthias Mrosek [107]. The individual components (intercooler, turbocharger, etc.) are modelled separately and are connected via storages to an overall mean value model. Engine test bed measurements are utilised to parametrise the sub models. The model structure can be regarded as semi-physical. The model states and inputs are shown in the sketched engine air path in Fig. A.1.



**Figure A.1:** Mean value model of the air path using a lumped parameter approach. Crossed out states in the storages are neglected. Fresh air enters through the air filter at the top left of the plot and the exhaust gas leaves the system at the lower left. The air path is modelled with low pressure egr (LP-EGR), a turbocharger with variable geometry turbine ( $s_t$ ) and a high pressure egr (HP-EGR). [107]

Model inputs are the ambient pressure  $p_a$ , the ambient temperature  $T_a$  and the actuator positions  $s$ . Model states are the temperatures  $T_i$ , the gas compositions  $x_i$  and the pressures  $p_i$  in the storages  $i$  and the turbocharger speed  $n_{tc}$ . The crossed out states are neglected to reduce the model complexity. The modelled air path states are regarded in more detail in the following.

The air path model covers the major dynamics of the Diesel engine. It also contains models for the actuator, whereas the actuator time constants are relatively small compared to the system time constants. These dynamics have however an effect on the optimisation results in Sect. 5.2. The system time constants are mainly introduced by the turbocharger and the egr paths. Further descriptions and discussions of this air path model can be found in [107, 109].

## A.1 Air path states

The applied states in the storages  $i$  are the temperatures  $T_i$ , the pressures  $p_i$  and the gas compositions  $x_i$ . The total gas mass in a storage  $m_i$  results from the ideal gas law,

$$p_i V_i = m_i R T_i. \quad (\text{A.1})$$

The air path states are calculated by mass and energy balances, see e. g. [43]. Regarding the air path, the engine is regarded as volumetric pump. The volumetric efficiency of this pump and the enthalpy of the exhaust, which is required to model the work of the turbocharger, are modelled by a look-up table identified from experiments [109, 185]. Flow equations through restrictions are modelled as valves or orifices, see e. g. [43].

### Gas composition and air to fuel ratio

The gas composition is defined as fraction of fresh air mass over the total gas mass

$$x = \frac{m_{\text{air, fresh}}}{m_{\text{gas, total}}}. \quad (\text{A.2})$$

$m_{\text{air, fresh}}$  is the fresh air mass with 20.9 % of oxygen and  $m_{\text{gas, total}}$  is the regarded total gas mass. The gas composition takes values between 0 and 1, standing for no fresh air and pure fresh air respectively. Especially the gas composition after combustion has an important influence on the emission formation. It is defined by the ratio of remaining fresh air after combustion and total gas mass in the cylinder. The remaining fresh air after combustion is given by the incoming fresh air mass in the cylinder  $m_{\text{air, eng}}$  minus the burnt air mass, utilised for combustion  $m_{\text{air, burnt}}$ . It is therefore defined as

$$x_{\text{eng, out}} = \frac{m_{\text{air, eng}} - m_{\text{air, burnt}}}{m_{\text{eng, in}} + m_{\text{Diesel}}}. \quad (\text{A.3})$$

The gas composition  $x_{2i}$  is the gas composition in the storage just before the cylinders. Multiplying this by the total gas mass in the cylinder  $m_{\text{eng, in}}$  gives the fresh air mass before combustion. Hence, eq. (A.3) is written as

$$x_{\text{eng, out}} = \frac{x_{2i} m_{\text{eng, in}} - u_{\text{inj}} \rho_{\text{Diesel}} L_{\text{st}}}{m_{\text{eng, in}} + u_{\text{inj}} \rho_{\text{Diesel}}}. \quad (\text{A.4})$$

The gas composition in a storage  $x_i$ , can directly be related to an air to fuel ratio  $\lambda_i$ . A commonly regarded air to fuel ratio is the air to fuel ratio after combustion, which is referred to as  $\lambda$ . It is defined as the total fresh air before the combustion over the required fresh air for combustion,

$$\lambda = \frac{m_{\text{air, eng}}}{u_{\text{inj}} \rho_{\text{Diesel}} L_{\text{st}}}. \quad (\text{A.5})$$

The required fresh air for combustion is calculated by the requested fuel mass  $u_{\text{inj}} \rho_{\text{Diesel}}$  multiplied by the stoichiometric air to fuel ratio  $L_{\text{st}}$ . Regarding the Diesel engine, the air to fuel ratio after combustion is always greater than one, since the Diesel engine requires a lean mixture.

The air to fuel ratio can also be defined for a storage  $i$ . Therefore, the air mass in a storage  $m_i$  is divided into fresh air mass and burnt gas mass in the storage by using the gas composition  $x_i$ ,

$$m_i = x_i m_i + (1 - x_i) m_i. \quad (\text{A.6})$$

The burnt gas mass can again be expressed by the fuel contained in the burnt gas  $m_{f,i}$  and the required air to burn this fuel,

$$(1 - x_i) m_i = m_{f,i} + m_{f,i} L_{st}. \quad (\text{A.7})$$

This can also be solved to

$$m_{f,i} = \frac{(1 - x_i) m_i}{1 + L_{st}}. \quad (\text{A.8})$$

The air to fuel ratio in a storage  $\lambda_i$  can then be defined as the fresh air in the storage  $x_i m_i$  plus the fresh air required to burn  $m_{f,i}$  over the fresh air required to burn  $m_{f,i}$ ,

$$\lambda_i = \frac{x_i m_i + m_{f,i} L_{st}}{m_{f,i} L_{st}} \quad (\text{A.9a})$$

$$= \frac{x_i m_i}{m_{f,i} L_{st}} + 1. \quad (\text{A.9b})$$

Then substituting eq. (A.8) into eq. (A.9b) leads to

$$\lambda_i = \frac{x_i m_i}{\frac{(1-x_i)m_i}{1+L_{st}} L_{st}} + 1 \quad (\text{A.10a})$$

$$= \frac{x_i(1 + L_{st})}{(1 - x_i)L_{st}} + \frac{(1 - x_i)L_{st}}{(1 - x_i)L_{st}} \quad (\text{A.10b})$$

$$= \frac{x_i + L_{st}}{(1 - x_i)L_{st}}. \quad (\text{A.10c})$$

Thus, the air fuel relates to the gas composition by

$$\lambda_i = \frac{x_i + L_{st}}{(1 - x_i)L_{st}}, \quad (\text{A.11})$$

or by solving this to  $x_i$ ,

$$x_i = \frac{L_{st}(\lambda_i - 1)}{L_{st}\lambda_i + 1}. \quad (\text{A.12})$$

This allows in particular, to relate the air to fuel ratio after combustion to the gas composition after combustion,

$$\lambda = \frac{x_{eng,out} + L_{st}}{(1 - x_{eng,out})L_{st}}, \quad (\text{A.13})$$

respectively

$$x_{eng,out} = \frac{L_{st}(\lambda - 1)}{L_{st}\lambda + 1}. \quad (\text{A.14})$$

There are *three* possibilities to determine the gas composition after combustion  $x_{eng,out}$ . The *first* is by measuring  $\lambda$  and applying eq. (A.14). Since the measurement of  $\lambda$  is located downstream of

the cylinder, this can only be applied for stationary conditions. For stationary conditions, the gas composition can also be calculated by the measured fresh air mass  $m_{\text{air}}$ ,

$$x_{\text{eng,out}} = \frac{m_{\text{air}} - u_{\text{inj}}\rho_{\text{Diesel}}L_{\text{st}}}{m_{\text{air}} + u_{\text{inj}}\rho_{\text{Diesel}}}. \quad (\text{A.15})$$

The *second* possibility to determine  $x_{\text{eng,out}}$  is by simulating the entire air path model. This is however sensitive to model uncertainties and the simulation of the air path model is computationally intensive for an implementation in the ECU.

The *third* possibility relates to the approach presented in [18] and applies the measured air to fuel ratio and eq. (A.12) to determine the gas composition of the recirculated exhaust gas,  $x_{\text{egr}}$  respectively  $x_{\text{lp-egr}}$  if available. Together with the modelled egr mass flow rates ( $m_{\text{egr}}$  and  $m_{\text{lp-egr}}$ ), the gas composition at intake is determined by

$$x_{2i} = \frac{m_{\text{air}} + x_{\text{egr}}m_{\text{egr}} + x_{\text{lp-egr}}m_{\text{lp-egr}}}{m_{\text{eng,in}}}, \quad (\text{A.16})$$

with  $m_{\text{eng,in}} = m_{\text{air}} + m_{\text{egr}} + m_{\text{lp-egr}}$ . Substituting this in eq. (A.4) gives the gas composition after combustion. Note that for a dynamic correct simulation, the dynamics between the egr storages and the intake storage, see Fig. A.1 need to be considered for eq. (A.16). Such a simulation requires however less computing time than a simulation of the entire air path model and is less sensitive to model errors, since the measured air to fuel ratio is applied.

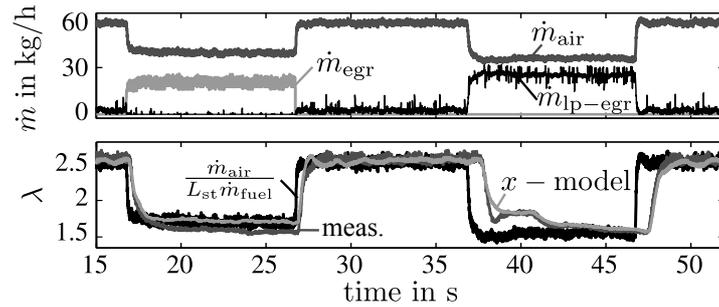
A quantity which is similar to the gas composition at intake is the egr rate

$$r_{\text{egr}} = \frac{m_{\text{egr}} + m_{\text{lp-egr}}}{m_{\text{eng,in}}}. \quad (\text{A.17})$$

This is a common applied variable for engine control. It has however the drawback that the quality of the exhaust is not considered, why some dynamical effects are not covered by the egr-rate which are covered by the gas composition. These *gas composition dynamics* are regarded in the following.

## A.2 Gas composition dynamics

The gas composition dynamics are dynamics in the quality of the exhaust gas. These dynamics are depicted for a step up and down in egr mass flow rate  $\dot{m}_{\text{egr}}$  and a step up and down in low-pressure egr mass flow rate  $\dot{m}_{\text{lp-egr}}$  in Fig. A.2. The upper plot shows the air mass flow rate and the modelled egr and low-pressure egr mass flow rates. The lower plot shows the stationary estimated air to fuel ratio (black line), which does not regard any dynamics, the measured air to fuel ratio (dark grey line) and the air to fuel ratio calculated by the air path model (light grey line). The differences between the stationary calculated estimation in black and the dynamic measurement in dark grey are due to the dynamics in the exhaust gas quality. Especially for the low-pressure egr with its longer travelling time, the differences are significant.



**Figure A.2:** Dynamics in the air to fuel ratio (lower plot) caused by a step in high-pressure egr mass flow rate  $\dot{m}_{\text{egr}}$  and low-pressure egr mass flow rate  $\dot{m}_{\text{lp-egr}}$  (upper plot). The black line shows the estimate considering no dynamics, the dark grey line is the measured  $\lambda$  and the light grey line is determined by the air path model. The dynamic variations in  $\lambda$  are due to the dynamics in the gas composition. [107].

At first, a step in high-pressure egr is performed,  $t = 17 - 27$  s, followed by a step in low-pressure egr,  $t = 37 - 47$  s. It can be seen that the stationary estimation (black line) reacts considerable fast to the steps. This dead time is relatively small for the high-pressure egr, but more significant for the low-pressure egr. This is because the measurement point of the air mass flow rate  $\dot{m}_{\text{air}}$ , which is applied for this estimation, is located at the beginning of the intake. It reacts therefore especially faster to the step in low pressure egr, since the mixing point of the low-pressure egr is just behind this measurement point, see Fig. A.1.

After the step in low-pressure egr mass flow rate at  $t = 37$  s, a first step in the measured  $\lambda$  is observable at  $t = 38$  s, which is due to the reduced fresh air mass flow rate. At  $t = 41$  s, there is a second step observable for the measured  $\lambda$ . This step is due to the gas composition dynamics and results from the travelling time of the egr. The first step lowers the gas composition after combustion. This lowered gas composition affects again the gas composition at intake, after being travelled through the exhaust pipe and the egr valve. The recirculated exhaust gas contains less oxygen, which further lowers the gas composition at intake. This change in gas composition at intake changes again the gas composition after combustion, which is observable as the second step in the measured air to fuel ratio  $\lambda$ . This second step is followed by another step, which is relatively small and only hardly observable.

These dynamics are covered by the air path model, which can be seen by the modelled  $\lambda$ , see  $x$ -model in Fig. A.2. Further discussions of these dynamics are given in [107, 108].

Hence, for a dynamical correct estimation of the gas composition, a model of the air path is necessary, as suggested by the second possibility in Sect. A.1. A less computational intensive possibility is given by the third possibility in Sect. A.1, which utilises the measured air to fuel ratio  $\lambda$  to estimation the gas compositions of the exhaust  $x_{\text{egr}}$  and  $x_{\text{lp-egr}}$ , see also [18].

## B Combustion Models Appendix

### B.1 Combustion model database

Depending on the purpose of a combustion model it might be necessary to apply different model inputs or different units for the applied model output. Therefore, a database of different models is stored from which an appropriate model can be selected, see Fig. B.1. The model number is shown in the first column, the second column states the applied model inputs and the third column the unit of the output. In the next columns the model quality for training and validation data is given. In the last two columns the number of applied parameters is stated.

### B.2 Models without in-cylinder pressure measurements

On the test bed, the engine is usually equipped with more sensors than for a series production. This is for example the case for in-cylinder pressure sensors. In-cylinder pressure sensors allow an analysis of the combustion process, but the sensors are relative expensive and the lifetime of a sensor is critical regarding the lifetime of an engine. Therefore, only few series configurations possess an in-cylinder pressure sensor. It might therefore be necessary to apply an alternative model input to the crank angle of 50 % mass fraction burnt  $\varphi_{Q50}$ , which is computed by in-cylinder pressure measurements. An alternative input is the crank angle of main injection  $\varphi_{mi}$ . Applying this inputs the model qualities as in Tab. B.1 are achieved.

**Table B.1:** Model qualities for NO<sub>x</sub>, soot and the engine torque using the crank angle of main injection  $\varphi_{mi}$  instead of the crank angle of 50% mass fraction burnt  $\varphi_{Q50}$ . Model qualities are stated for the training and validation data as in Sect. 3.3.

Model	Training		Validation		# parameters
	$R^2$	RMSE	$R^2$	RMSE	
$\dot{m}_{nox}$	0.995	1.51 mg/s	0.987	2.18 mg/s	427
$\dot{m}_{soot}$	0.952	0.06 mg/s	0.876	0.07 mg/s	451
$M_{eng}$	0.997	1.54 Nm	0.980	2.90 Nm	261
$c_{nox}$	0.991	25.4 ppm	0.976	37.1 ppm	412
$c_{soot}$	0.917	1.02 mg/m <sup>3</sup>	0.840	1.28 mg/m <sup>3</sup>	435

The crank angle of 50 % mass fraction burnt  $\varphi_{Q50}$  can also be modelled. Therefore, the global-local model structure, as is applied for the combustion model, is utilised. The inputs are  $n_{eng}$ ,  $u_{inj}$ ,  $m_{air}$ ,  $p_{2i}$ ,  $T_{2i}$  and  $\varphi_{mi}$  and the model quality is stated in Tab. B.2.

Emission Models NO <sub>x</sub>								
No.	inputs	unit (output)	Training		Validation		Parameter	
			R2	RMSE	R2	RMSE	total	effective
1	[ mAir, p_2, phi_Q50, T_2 ]	ppm	0,992	23,05	0,979	34,66	456	427,3
2	[ x_{eng-out}, p_2, phi_Q50, T_2 ]	ppm	0,988	29,01	0,947	55,63	404	378,5
3	[ mAir, p_2, phi_Q50, r_egr ]	ppm	0,992	22,83	0,980	33,95	414	391,5
4	[ x_{eng-out}, p_2, phi_Q50, r_egr ]	ppm	0,990	26,45	0,970	41,92	435	372,7
5	[ mAir, p_2, phi_MI, T_2 ]	ppm	0,991	25,37	0,976	37,06	412	386,7
6	[ x_{eng-out}, p_2, phi_MI, T_2 ]	ppm	0,986	30,68	0,943	57,67	414	367,1
7	[ mAir, p_2, phi_MI, r_egr ]	ppm	0,992	23,77	0,979	34,80	399	382,2
8	[ x_{eng-out}, p_2, phi_MI, r_egr ]	ppm	0,989	27,64	0,971	41,38	424	365,5
9	[ mAir, p_2, phi_Q50, T_2 ]	mg/s	0,996	1,45	0,988	2,09	424	396,5
10	[ x_{eng-out}, p_2, phi_Q50, T_2 ]	mg/s	0,990	2,18	0,956	4,01	396	353,3
11	[ mAir, p_2, phi_Q50, r_egr ]	mg/s	0,997	1,27	0,989	2,01	410	386,1
12	[ x_{eng-out}, p_2, phi_Q50, r_egr ]	mg/s	0,994	1,73	0,977	2,89	426	361,8
13	[ mAir, p_2, phi_MI, T_2 ]	mg/s	0,995	1,51	0,987	2,18	427	403,7
14	[ x_{eng-out}, p_2, phi_MI, T_2 ]	mg/s	0,990	2,17	0,958	3,93	403	357,2
15	[ mAir, p_2, phi_MI, r_egr ]	mg/s	0,997	1,28	0,989	2,05	415	393,4
16	[ x_{eng-out}, p_2, phi_MI, r_egr ]	mg/s	0,994	1,71	0,979	2,78	416	357,7
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

**Figure B.1:** Database of the NO<sub>x</sub> emission model. The various model structures apply different model inputs and model different units of the regarded output. The database can be utilised to select the suited model for an application.

The modelled crank angle of 50 % mass fraction burnt  $\varphi_{Q50}$  can then be applied as input for the combustion models. The qualities of these models are stated in Tab. B.3. The qualities are similar to the results as in Tab. B.1, but the soot model  $c_{soot}$  is superior applying the modelled  $\varphi_{Q50}$ .

### B.3 Computing the global-local model structure

To illustrate the working principle of the global-local model structure, the outputs of the local NO<sub>x</sub> models are presented by the matrix plot in Fig. B.2. Each of the 21 plots belongs to one local model, compare to Fig. 3.4. The output of a local model is shown by the black line and the current value of the local weighting function, indicating the validity of a local model, by the grey line.

**Table B.2:** Model qualities for training and validation data for the crank angle of 50% mass fraction burnt  $\varphi_{Q50}$ .

Model	Training		Validation		# parameters
	$R^2$	RMSE	$R^2$	RMSE	
$\varphi_{Q50}$	0.985	0.45 °CA	0.959	0.75 °CA	385

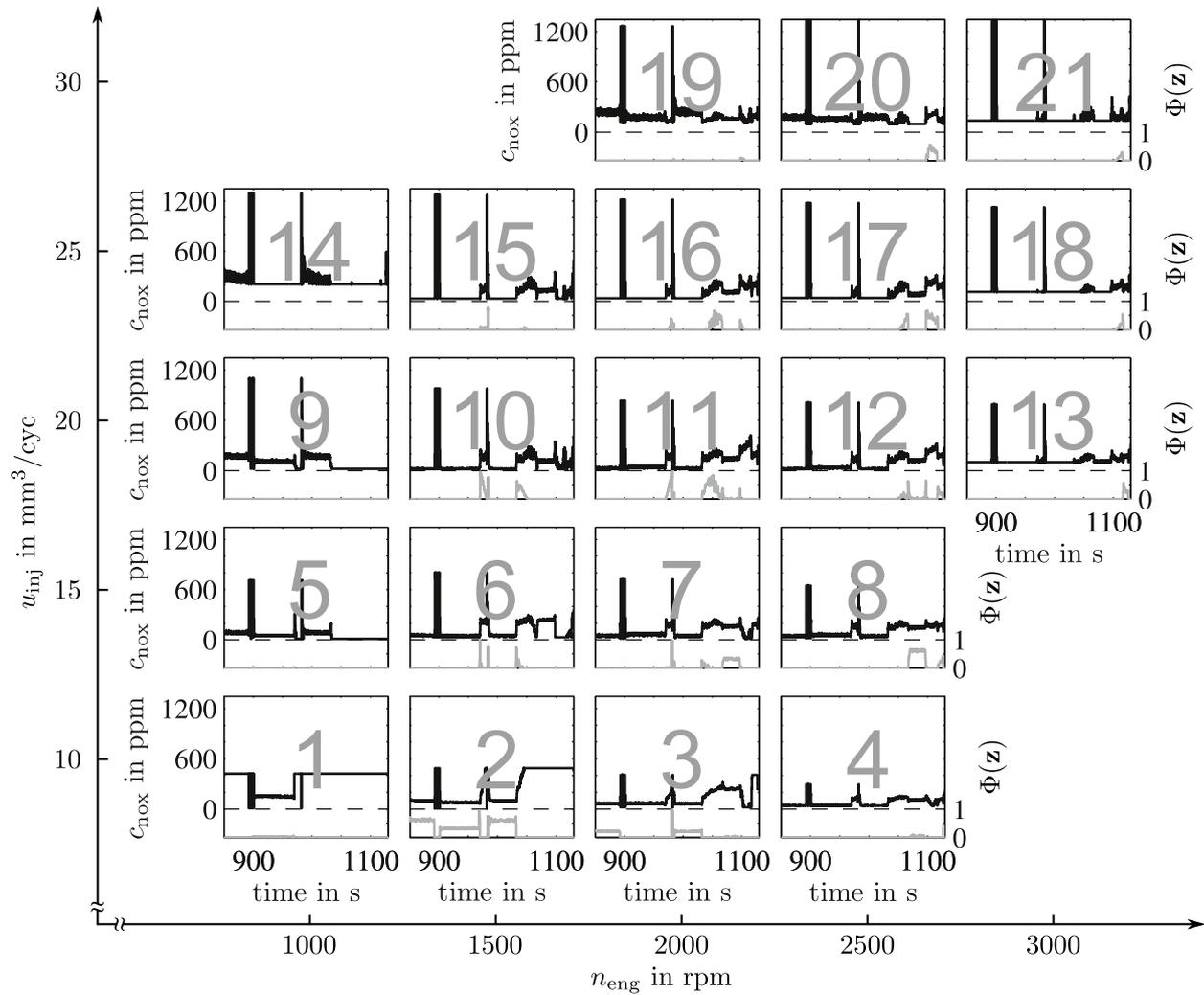
**Table B.3:** Model qualities for  $\text{NO}_x$ , soot and the engine torque using the modelled crank angle of 50% mass fraction burnt instead of the calculated based on incylinder pressure measurements. Model qualities are stated for training and validation data.

Model	Training		Validation		# parameters
	$R^2$	RMSE	$R^2$	RMSE	
$\dot{m}_{\text{nox}}$	0.995	1.50 mg/s	0.986	2.27 mg/s	435
$\dot{m}_{\text{soot}}$	0.956	0.05 mg/s	0.886	0.07 mg/s	435
$M_{\text{eng}}$	0.997	1.52 Nm	0.981	2.85 Nm	240
$c_{\text{nox}}$	0.991	24.8 ppm	0.976	37.7 ppm	428
$c_{\text{soot}}$	0.928	0.95 mg/m <sup>3</sup>	0.891	1.05 mg/m <sup>3</sup>	430

The presented data is from the extra-urban part of the NEDC. Looking at the local weighting functions in grey, it can be seen that for example model 2 ( $n_{\text{eng}} = 1500$  rpm,  $u_{\text{inj}} = 10$  mm<sup>3</sup>/cyc) is valid until  $t = 900$  s together with model 3 ( $n_{\text{eng}} = 2000$  rpm,  $u_{\text{inj}} = 10$  mm<sup>3</sup>/cyc). Thereafter, model 2 is valid with the not shown model for  $n_{\text{eng}} = 1500$  rpm and  $u_{\text{inj}} = 5$  mm<sup>3</sup>/cyc, before model 2 is again valid with model 3 until  $t = 1030$  s. Thereafter, other local models in the engine operation range are valid. For the time  $t > 1050$  s, model 2 is not valid and operates in extrapolation region, which can be seen on the limited local model output. These limitations avoid unreasonable for extrapolation regions of local models.

## B.4 Dynamic emission characteristics

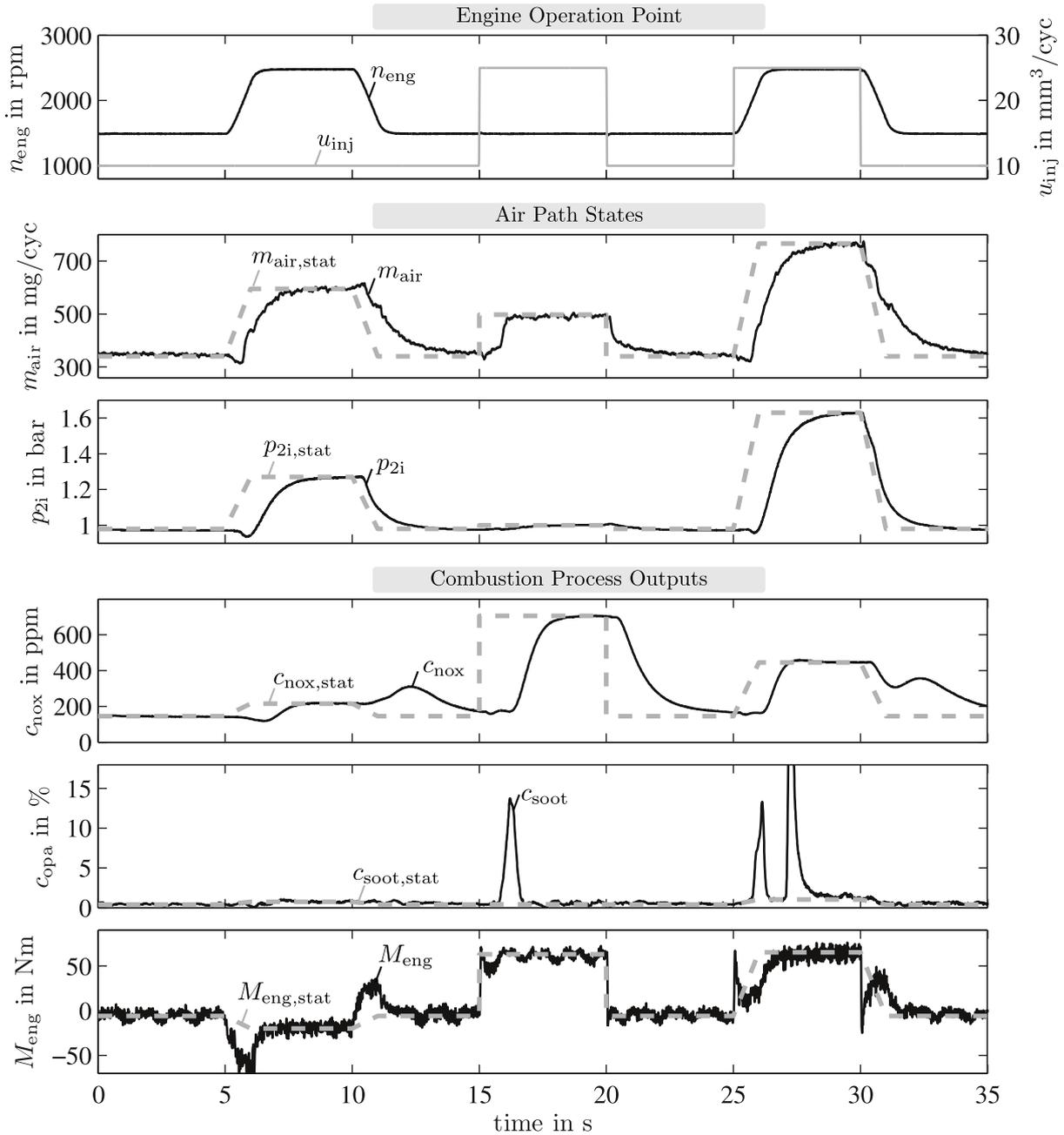
The dynamic emission characteristics are investigated in Sect. 3.6 by a simulation of the dynamic air path model connected to the stationary combustion model. In Fig. 3.13 three typical transients are regarded, which are a step up and down in injection quantity, a ramp up and down in engine speed and a step up and down in injection quantity each followed by a ramp up respectively down in engine speed. The latter two transients can be regarded as acceleration and deceleration event. The same transients, with different set points for the air path states, are investigated by measurements from the engine test bed in Fig. B.3. The engine trajectory is shown in the topmost plot. The air path states are shown in the two plots below and the lower three plots show the combustion outputs.



**Figure B.2:** Local model outputs of the  $\text{NO}_x$  model showing the 21 local models over the engine operation points. Model outputs are shown by the black lines and the validity of the local models  $\Phi(\mathbf{z})$  by the grey lines. The shown time window is from the extra urban part of the NEDC.

The predicted combustion outputs are similar to the simulations in Fig. 3.13. The overshoots in  $\text{NO}_x$  emissions are observable for the ramps down in engine speed. These overshoots are due to the increased air mass per cycle  $m_{\text{air}}$ , which is however less pronounced for the measured  $m_{\text{air}}$  since the measurement is at the beginning of the intake, see Fig. A.1, which smooths the measured overshoot in air mass per cycle.

The soot peaks are shown by the opacity measurements for the steps up in injection quantity. The dead time of the measurement dynamics is approximately 1 s. The two peaks for the second step up in injection quantity are probably due to a protective mechanism of the Opacimeter because of the high soot emissions. The dynamics in torque measurement are mainly due to the engine speed control of the asynchronous machine. For an engine ramp up, the torque of the asynchronous machine is reduced such that the engine accelerates. This reduced torque is then observable in the



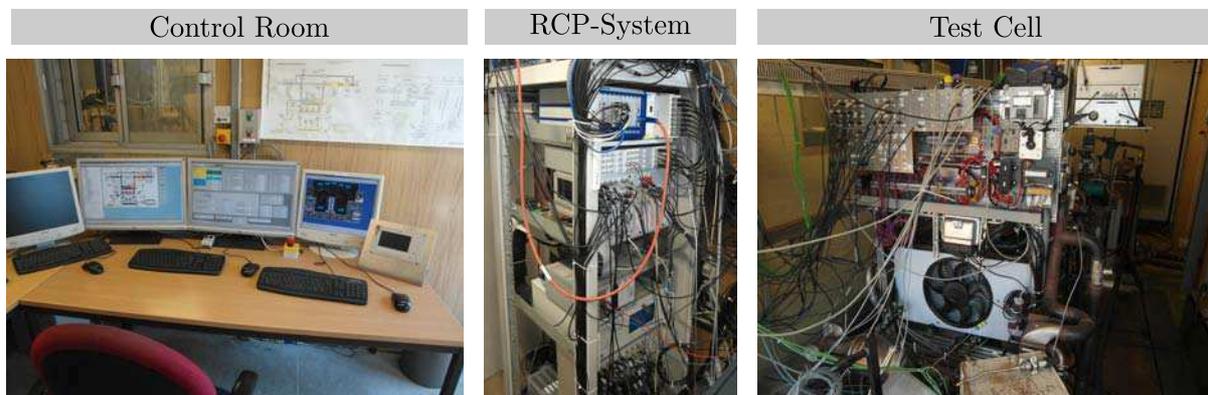
**Figure B.3:** Measured combustion outputs for a dynamic engine operation. Engine operation point trajectory is shown in the topmost plot. The first two transients illustrate separately the dynamics in engine speed and injection quantity, whereas the latter can be regarded as typical acceleration respectively deceleration event. The model actuators are adjusted by open loop controls, which is why there are relatively long settling times for  $m_{air}$  and  $p_{2i}$  (second and third plot). Measured combustion outputs are shown in the lower three plots.

---

measurements, see  $t = 5 - 6$  s and  $t = 25 - 26$  s. The contrary can be observed for ramps down in engine speed, see  $t = 10 - 11$  s and  $t = 30 - 31$  s.

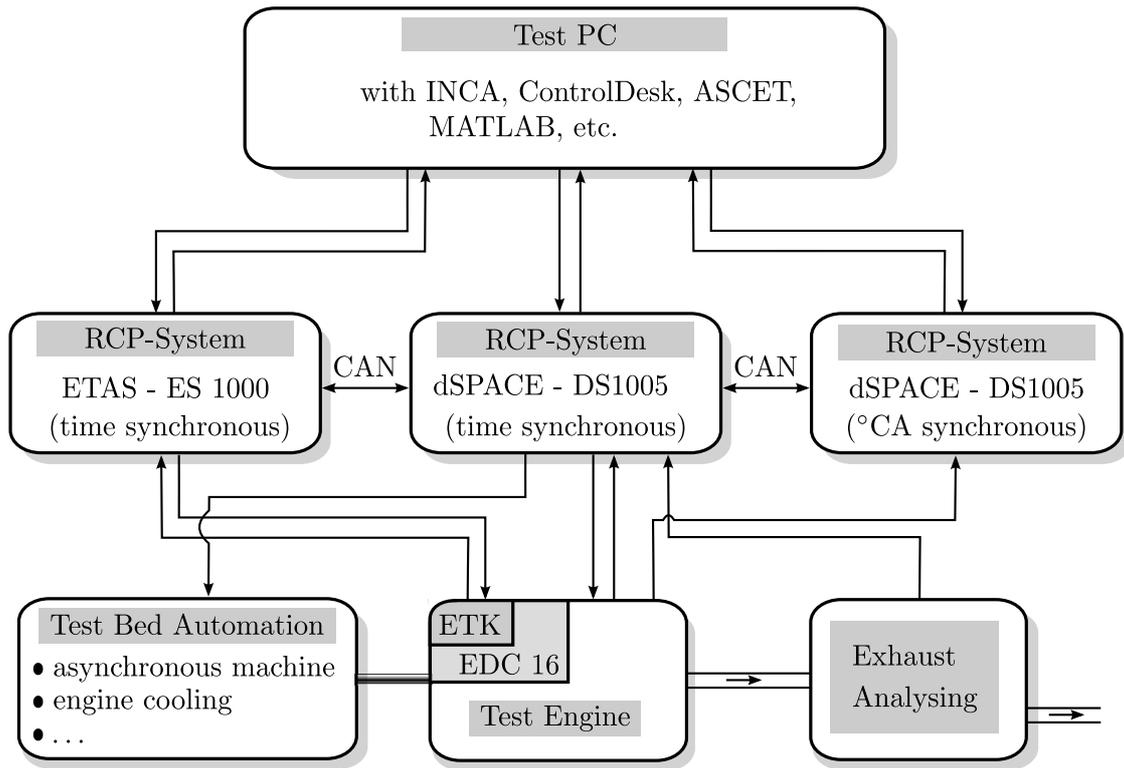
## C Engine Test Bed

The engine test bed at the Institute of Automatic Control and Mechatronics at the TU Darmstadt is shown in Fig. C.1. It is a high dynamical test bed on which measurements of the engine operation are recorded and new control functions are designed. Control functions can in general be categorised into control functions for the test bed automation and control functions for the engine control. The test bed automation mainly includes the control of the engine speed by the asynchronous machine and the control of the engine temperature. The asynchronous machine has a power of 160 kW, a maximum torque of 300 Nm and requires less than 5 ms to adjust a variation in torque. The test engine is mounted on a roller carriage such that various engines can be tested on the test bed. With this roller carriage and the applied plug connections, engines can be changed in a few hours.



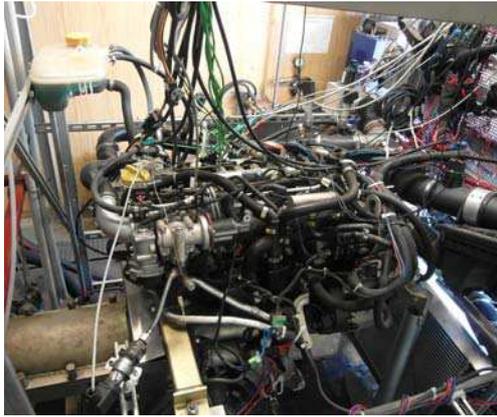
**Figure C.1:** Photos of the test bed at the Institute of Automatic Control and Mechatronics at the TU Darmstadt. Left picture shows the control room with the test PC, middle picture shows the RCP-Systems and the right picture the test cell with the test engine and the panel to switch between ECU and bypass control of actuators.

The control functions for the test engine can either be realised via the ETK of the ECU or by bypassing the ECU with the RCP-System. The former varies the set points in the ECU and the actuators are still controlled by the ECU, whereas the actuators are directly controlled by the RCP-System for the latter. A manual switch, shown in the upper left of the right picture in Fig. C.1, enables to select between those possibilities. Besides manipulating the actuators, the RCP-system enables to record measurements. Measurements are given for series sensors, which are read from the ECU, and also for additional sensors, which are directly connected to the RCP-system. The latter are recorded either time- or crank angle-synchronous. Crank angle-synchronous measurements are given among others for the in-cylinder pressure sensors with a resolution of  $1^\circ\text{CA}$ . With these measurements, combustion characteristics, such as the crank angle of 50 % mass fraction burnt  $\varphi_{Q50}$  are calculated. The real time computation of these characteristics enables the closed loop control of these on the test bed [83]. A schematic overview of the signal flows on the test bed is shown in Fig. C.2.



**Figure C.2:** Test bed automation at the Institute of Automatic Control and Mechatronics at the TU Darmstadt. The test PC is the interface to the test bed engineer. Rapid Control Prototyping (RCP) systems are utilised to manipulate the engine control and to record the measurements. The engine control is either manipulated by changes via the ETK of the ECU or by using the bypass and controlling the actuators directly with the RCP-System. Signal flows are indicated as arrows.

The utilised test engine is a Opel/Fiat 1.9l Common-Rail Diesel engine. It is equipped with a high-pressure egr system and a turbocharger with variable geometry turbine. An additional low-pressure egr system is retrofitted to the engine. A picture of the engine and a data sheet with the main engine characteristics is shown in Fig. C.3. The picture shows various sensor connections and the drive shaft to the asynchronous machine in the lower left. Besides the series sensors, there are additional sensors for temperatures, pressures, actuator positions, the engine torque and the exhaust gas. Exhaust gas measurements are given by a lambda sensor by Bosch, a NO<sub>x</sub> sensor by NGK, a Micro Soot sensor by AVL and an Opacimeter by AVL. The soot measurement correlates to the opacity measurement, why these can be used for equivalent investigations of the engine operation. The opacity measurement is therefore applied to validate the model-based smoke limitation, since the Micro Soot sensor is not available for these measurements.



---

Manufacturer	GM/Opel and Fiat
Engine designation	Z19DTH (4V)
Combustion method	Four-stroke Diesel
Number of cylinders	4
Displacement	1910 cm <sup>3</sup>
Compression ratio	17.5
max power (at engine speed)	110 kW (4000 rpm)
max torque (at engine speed)	315 Nm (2000 rpm)
idle speed	850 rpm
max speed	5100 rpm

---

**Figure C.3:** Left picture shows the test engine with various sensor connections and the drive shaft to the asynchronous machine at the lower left. The main engine characteristics of the test engine are shown in the table on the right.

## D Mathematical Appendix

### D.1 Recursive least squares algorithm

The recursive least squares solution for a linear least squares problem, see eq. (2.8), is given by [67]

$$\hat{\mathbf{w}}(k+1) = \hat{\mathbf{w}}(k) + \gamma(k+1) (y(k+1) - \mathbf{x}^T(k+1)\hat{\mathbf{w}}(k)) \quad (\text{D.1})$$

with

$$\gamma(k+1) = \frac{1}{\alpha + \mathbf{x}^T(k+1)\mathbf{P}(k)\mathbf{x}(k+1)} \mathbf{P}(k)\mathbf{x}(k+1) \quad (\text{D.2})$$

$$\mathbf{P}(k+1) = \frac{1}{\alpha} (\mathbf{I} - \gamma(k+1)\mathbf{x}^T(k+1)) \mathbf{P}(k) \quad (\text{D.3})$$

and the input vector consisting of the grid point weights at the discrete time  $k$

$$\mathbf{x}(k)^T = \left( \Phi_1(\mathbf{u}(k)) \quad \Phi_2(\mathbf{u}(k)) \quad \dots \quad \Phi_L(\mathbf{u}(k)) \right). \quad (\text{D.4})$$

For an calculation of the recursive least square solution, initial values need to be chosen for the parameter vector  $\mathbf{w}(0)$  and the matrix  $\mathbf{P}(0)$ . If a least squares solution as in eq. (2.8) exists before an adaptation by the recursive algorithm is applied, the initial values can be given as

$$\hat{\mathbf{w}}(0) = \hat{\mathbf{w}} \quad (\text{D.5})$$

and

$$\mathbf{P}(0) = (\mathbf{X}^T\mathbf{X})^{-1}. \quad (\text{D.6})$$

If no a-priori information is given, the initial values are chosen as

$$\hat{\mathbf{w}}(0) = \mathbf{0} \quad (\text{D.7})$$

and

$$\mathbf{P}(0) = \beta \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}; \quad \beta = 100 \dots 10000 \quad (\text{D.8})$$

With a selection as in eq. (D.8), the influence of the initial value  $\mathbf{w}(0)$  vanishes with increasing measurement time. Independent from the initial values, a forgetting factor  $\alpha \leq 1$  has to be chosen. A value  $\alpha = 1$  corresponds to no forgetting and  $\alpha < 1$  enables an adaption to time variant processes. The corresponding forgetting factor needs then to be chosen with regard to the alteration. A forgetting factor of  $\alpha > 0.9$  is mostly sufficient.

## D.2 Regularisation for look-up tables

The regularisation in eq. (2.15)

$$\hat{\mathbf{w}}_{\text{reg}} = (\mathbf{X}^T \mathbf{X} + \lambda^2 \mathbf{\Gamma}^T \mathbf{\Gamma})^{-1} \mathbf{X}^T \mathbf{y}, \quad (\text{D.9})$$

with eq. (2.12)

$$\lambda \mathbf{\Gamma} = \lambda \begin{pmatrix} -2 & 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 & 0 & 0 & \dots \\ 1 & -3 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 & 0 & 0 & \dots \\ 0 & 1 & -3 & 1 & \dots & 0 & 0 & 1 & \dots & 0 & 0 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \dots & \ddots & \ddots & \ddots & \dots & \ddots & \ddots & \ddots & \dots \\ 0 & 1 & 0 & 0 & \dots & 1 & -4 & 1 & \dots & 0 & 1 & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots & \dots & \ddots & \ddots & \ddots & \dots & \ddots & \ddots & \ddots & \dots \end{pmatrix} \quad (\text{D.10})$$

penalises the second order derivative for interior grid points and the first order derivative for grid points on the boundary. This can be seen by looking at the difference quotients, which are given in the case of the second order derivatives by

$$\frac{\partial^2 w_{i,j}}{\partial u_1^2} = \frac{w_{i-1,j} - 2w_{i,j} + w_{i+1,j}}{\Delta u_1^2} \quad (\text{D.11a})$$

$$\frac{\partial^2 w_{i,j}}{\partial u_2^2} = \frac{w_{i,j-1} - 2w_{i,j} + w_{i,j+1}}{\Delta u_2^2}. \quad (\text{D.11b})$$

Applying these on an interior point, a row such as the lower row in eq. (D.10) results. The denominator in eq. (D.11) is not of interest, since the equations are set to zero. The denominator is however of interest if non-equidistant grid point distributions are applied.

Assuming that the discretisation is given by  $\Delta_i(u_{i,j}) = u_{i+1,j} - u_{i,j}$  for the first dimension and by  $\Delta_j(u_{i,j}) = u_{i,j+1} - u_{i,j}$  for the second dimension, then the difference quotients of the second order derivatives are derived from the Taylor series to

$$\frac{\partial^2 w_{i,j}}{\partial u_1^2} = \frac{2}{\Delta_i(u_{i-1,j}) + \Delta_i(u_{i,j})} \left( \frac{w_{i-1,j} - w_{i,j}}{\Delta_i(u_{i-1,j})} + \frac{w_{i+1,j} - w_{i,j}}{\Delta_i(u_{i,j})} \right) \quad (\text{D.12a})$$

$$\frac{\partial^2 w_{i,j}}{\partial u_2^2} = \frac{2}{\Delta_j(u_{i,j-1}) + \Delta_j(u_{i,j})} \left( \frac{w_{i,j-1} - w_{i,j}}{\Delta_j(u_{i,j-1})} + \frac{w_{i,j+1} - w_{i,j}}{\Delta_j(u_{i,j})} \right). \quad (\text{D.12b})$$

Setting eq. (D.12a) and eq. (D.12b) to zero it can be seen that the discretisations have an influence, why these need to be regarded for non-equidistant grid point distributions. Thus, the non-equidistant distribution influences the elements in  $\mathbf{\Gamma}$ .

Since the difference quotients of the second order derivatives in eq. (D.11) can not be calculated on the boundary of the lattice, the first order derivative is penalised there. The difference quotients for the first order derivatives on the boundaries of the lower sides are

$$\frac{\partial w_{i,j}}{\partial u_1} = \frac{w_{i+1,j} - w_{i,j}}{\Delta u_1} \quad (\text{D.13a})$$

$$\frac{\partial w_{i,j}}{\partial u_2} = \frac{w_{i,j+1} - w_{i,j}}{\Delta u_2}, \quad (\text{D.13b})$$

and for the boundaries of the upper sides

$$\frac{\partial w_{i,j}}{\partial u_1} = \frac{w_{i,j} - w_{i-1,j}}{\Delta u_1} \quad (\text{D.14a})$$

$$\frac{\partial w_{i,j}}{\partial u_2} = \frac{w_{i,j} - w_{i,j-1}}{\Delta u_2}. \quad (\text{D.14b})$$

The difference quotients for the first order derivatives are not affected by a non-equidistant distribution. With the difference quotients given for the second and the first order derivatives, the matrix  $\Gamma$  in eq. (D.10) results.

### Dimension depending regularisation

It is further possible to use different smoothing parameters in each dimension,  $\lambda_1$  and  $\lambda_2$  respectively. Then the regularisation for an interior point changes to

$$\lambda \Gamma = \begin{pmatrix} \dots & \dots \\ 0 & \lambda_2 & 0 & 0 & \dots & \lambda_1 & -2\lambda_1 - 2\lambda_2 & \lambda_1 & \dots & 0 & \lambda_2 & 0 & \dots & \dots \\ \dots & \dots \end{pmatrix}. \quad (\text{D.15})$$

The other rows are defined analogue.

### Regularisation for the recursive least squares

The recursive least squares algorithm as presented in App. D.1 enables the adaption of a look-up table to time variant processes. However, with an ongoing adaption also the smoothness of the look-up table vanishes. To maintain the smoothness of a look-up table, a regularisation needs to be performed during the recursive algorithm. The regularisation can either be performed after every incoming measurement or after several measurements.

For a recursive adaption of the parameters, eq. (D.1) is evaluated for the new measurement. Then, the regularisation need to be calculated. Therefore, eq. (D.10) needs to be incorporated into eq. (D.1). This needs to be done line by line such that for every measurement  $L = L_1 L_2$  further evaluations of eq. (D.1) are necessary for model regularisation.

This increases the computational effort significantly. It is therefore superior to apply the smoothing not after every recursive step, but after several measurements. Than the regularisation parameter needs to be calculated recursively by

$$\lambda = \tilde{\lambda} f_\lambda(k+1) \quad (\text{D.16})$$

with

$$f_\lambda^2(k+1) = f_\lambda^2(k) + \left( \frac{\mathbf{x}(k+1)^T \mathbf{x}(k+1)}{\text{trace}(\mathbf{\Gamma}^T \mathbf{\Gamma})} \right). \quad (\text{D.17})$$

The regularisation parameter  $\tilde{\lambda}$  is the applied regularisation parameter for the non-recursive solution. The regularisation can then be performed after an arbitrary number of recursive steps. After a

regularisation at discrete time  $k$ , the  $f_\lambda$  is set to zero,  $f_\lambda(k) = 0$ . A detailed discussion about regularisations of look-up tables with the recursive least squares algorithm and a derivation of eq. (D.17) is given in [75].

### D.3 Pyramidal weighting function

The pyramidal weighting function is given as

$$\Phi_j(\mathbf{z}) = \frac{\prod_{i=1}^{n_z} 1 - \frac{\Delta z_i}{d_{j,i}}}{\sum_{j=1}^M \prod_{i=1}^{n_z} 1 - \frac{\Delta z_i}{d_{j,i}}}. \quad (\text{D.18})$$

with

$$\Delta z_i = \max(|z_i - z_{0,j,i}|, 0). \quad (\text{D.19})$$

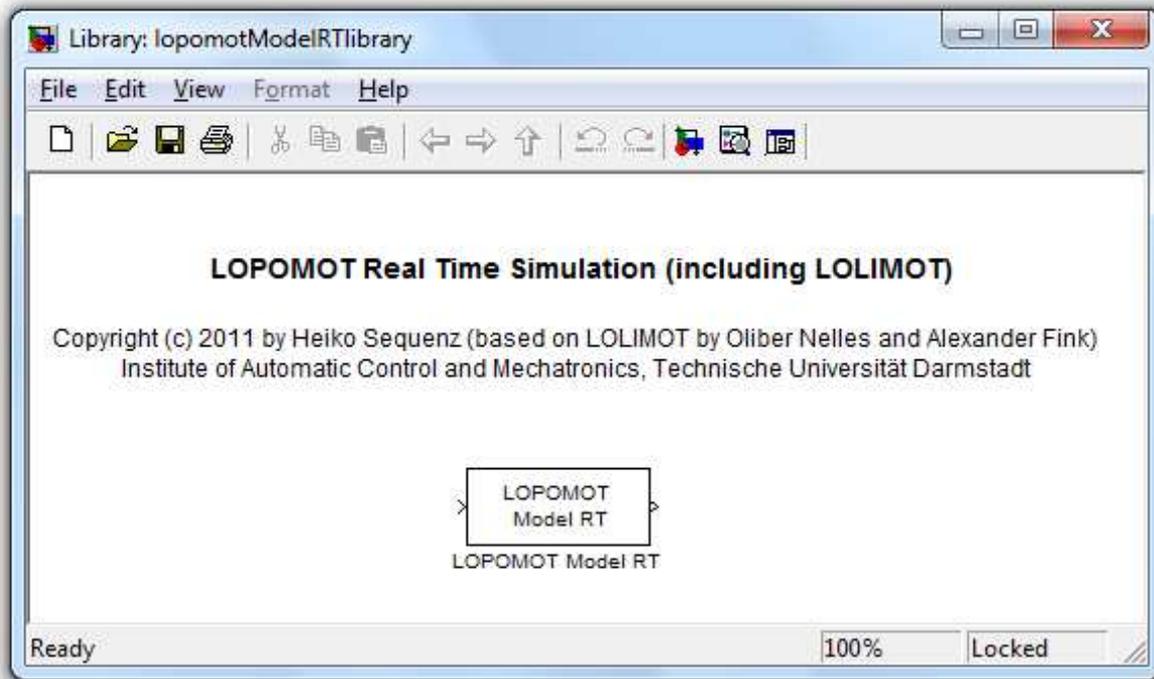
$z_{0,j,i}$  is the  $i$ -th dimension of the  $j$ -th model centre and  $d_{j,i}$  is the distance to next model centre in the  $i$ -th dimension. The pyramidal weighting function has the advantage to the Gaussian weighting function in eq. (2.48) that it has a compact support and that it simplifies to a linear interpolation if the model centres are distributed on an equidistant lattice. This is the case for the emission models presented in Chap. 3, which eases the calculation of the structure regarding the implementation in an ECU, see Sect. 6.1.

### D.4 LOPOMOT c-function

The LOPOMOT model is also programmed as c-function for a fast model simulation. With this implementation and the corresponding simulink block, see Fig. D.1, a LOPOMOT model can easily be integrated in block-oriented programming. It is further possible to compile such a model for a rapid control prototyping system.

### D.5 Derivation of the weights for the multi-objective optimisation function

The emission limits for  $\text{NO}_x$  and soot are of a different magnitude. Therefore, a uniform distribution of the weights  $k_{\text{nox}}$  and  $k_{\text{soot}}$  would not show a satisfactory distribution on the Pareto front. The weights are therefore distributed based on three distinguished values, which are given from the legislative limits. The EU 5 emission limits are  $\text{NO}_{x,\text{limit}} = 180 \text{ mg/km}$  and  $\text{soot}_{\text{limit}} = 5 \text{ mg/km}$ . For an equal weighting, the ratio  $k_{\text{nox}} : k_{\text{soot}}$  should therefore be  $180 : 5$  which is identical to  $36 : 1$ . A uniform distribution in  $\tilde{k}_{\text{nox}}$  should therefore fulfil the following three conditions:



**Figure D.1:** Simulink block of a LOPOMOT model. The simulink block is programmed as c-function and can be used for a fast simulation of a LOPOMOT model. It can further be directly compiled on a rapid control prototyping system for a block-oriented system design.

- $\tilde{k}_{\text{nox}} = 0$  corresponds to a weighting  $k_{\text{nox}} = 0 \rightarrow \{0, 0\}$
- $\tilde{k}_{\text{nox}} = 0.5$  corresponds to a weighting  $k_{\text{nox}} = \frac{1}{37} \rightarrow \{0.5, \frac{1}{37}\}$
- $\tilde{k}_{\text{nox}} = 1$  corresponds to a weighting  $k_{\text{nox}} = 1 \rightarrow \{1, 1\}$

With these three points a function is identified to describe the distribution of weights. A suited model structure for this is the exponential function,

$$k_{\text{nox}} = a \cdot \exp(b \cdot \tilde{k}_{\text{nox}}) - c. \quad (\text{D.20})$$

Identifying a, b and c with the three distinguished values  $\{0, 0\}$ ,  $\{0.5, \frac{1}{37}\}$  and  $\{1, 1\}$ , results in

$$k_{\text{nox}} = \frac{(36^{2\tilde{k}_{\text{nox}}})}{35 \cdot 37}. \quad (\text{D.21})$$

# Bibliography

- [1] K. Ackerl, H. Hochschwarzer, M. Paulweber, W. Kriegler, and L. Bloder. Vollautomatische Motorkennfeldoptimierung. In *13. Internationales Wiener Motorensymposium*, pages 245–271, Wien, Austria, 1992.
- [2] L. A. Aguirre and S. A. Billings. Improved structure selection for nonlinear models based on term clustering. *International Journal of Control*, 62(3):569–587, 1995.
- [3] D. Ajtay and M. Weilenmann. Compensation of the exhaust gas transport dynamics for accurate instantaneous emission measurements. *Environmental science & technology*, 38(19):5141–5148, 2004.
- [4] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.
- [5] D. Alberer and L. del Re. Fast Oxygen Based Transient Diesel Engine Operation. In *SAE World Congress and Exhibition*, number 2009-01-0622, pages 405–413, Detroit, USA, 2009.
- [6] D. Alberer and L. del Re. Optimization of the transient Diesel engine operation. In *9th International Conference on Engines and Vehicles*, number 2009-24-0113, Capri, Naples, Italy, 2009.
- [7] D. Alberer, M. Hirsch, and L. del Re. Optimal Selection of Control Inputs for Diesel Engines. In *17th IFAC World Congress*, pages 10704–10709, Seoul, South Korea, 2008.
- [8] E. Alfieri. *Emissions-Controlled Diesel Engine*. Dissertation, Swiss Federal Institute of Technology Zürich, 2009.
- [9] M. Ammann. *Modellbasierte Regelung des Ladedrucks und der Abgasrückführung beim aufgeladenen PKW-Common-Rail-Dieselmotor*. Dissertation, ETH Zürich, 2003.
- [10] K.J. Åström and T. Hägglund. *Advanced PID Control*. ISA - The Instrumentation, Systems, and Automation Society, North Carolina, USA, 2006.
- [11] S.J. Atiye and N.A. Babak. PiLiMoT: A Modified Combination of LoLiMoT and PLN Learning Algorithms for Local Linear Neurofuzzy Modeling. *Journal of Control Science and Engineering*, 2011:1–9, 2011.
- [12] O. Bänfer and O. Nelles. Polynomial model tree (POLYMOT) - A new training algorithm for local model networks with higher degree polynomials. In *7th IEEE International Conference on Control and Automation*, pages 1571–1576, Christchurch, New Zealand, 2009.
- [13] W. Baumann, K. Klug, B.U. Köhler, and K. Röpke. Modeling of transient diesel engine emissions. In *Design of Experiments in Engine Development V*, number 5, Berlin, Germany, 2009.

- [14] W. Baumann, K. Röpke, S. Stelzer, and A. Frank. Model-Based Calibration Process for Diesel Engines. In *4th International Symposium on Development Methodology*, pages 196–208, Wiesbaden, Germany, 2011.
- [15] R.E. Bellman. *Adaptive Control Processes: A Guided Tour*. Princeton University Press, New Jersey, USA, 1961.
- [16] M. Benz. *Model-Based Optimal Emission Control of Diesel Engines*. Dissertation, ETH Zürich, 2010.
- [17] M. Benz, C.H. Onder, and L. Guzzella. Engine Emission Modeling Using a Mixed Physics and Regression Approach. *Journal of Engineering for Gas Turbines and Power*, 132(4):042803, 2010.
- [18] C. Bessai, E. Stölting, and R. Gratzke. Virtual Oxygen Sensor in the Intake Manifold of a Diesel Engine. *MTZ - worldwide*, 72:34–39, 2011.
- [19] A. Bittermann, E. Kranawetter, J. Krenn, B. Ladein, T. Ebner, H. Altenstrasser, H.-M. Koegeler, and K. Gschweidl. Emissions Development of Vehicle Diesel Engines by Means of DoE and Computer Simulation. *MTZ - worldwide*, 06:15–19, 2004.
- [20] D. Böhme, R. Isermann, M. Martiny, J. Kern, and M. Kerkau. Quasi-stationary measurement method for the ECU-functions of the cylinder charge determination and the torque model. In *3rd International Symposium on Development Methodology*, pages 84–96, Wiesbaden, Germany, 2009.
- [21] A. Brahma, A. Serrani, and G. Rizzoni. Modeling, Identification and State Estimation of Torque and NO<sub>x</sub> Dynamics in Common-Rail Diesel Engines. In *4th IFAC Advances in Automotive Control*, pages 251–256, Salerno, Italy, 2004.
- [22] J. Bredenbeck. Statistische Versuchplanung für die Online-Optimierung von Verbrennungsmotoren. *MTZ - Motortechnische Zeitschrift*, 60(11):740–744, 1999.
- [23] I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Muhlig. *Handbook of Mathematics*. Verlag Harri Deutsch, Frankfurt am Main, Germany, 2000.
- [24] M. Büchel and M. Thomas. Roll out of a fast calibration approach for engine base calibration. In *3rd International Symposium on Development Methodology*, pages 256–271, Wiesbaden, Germany, 2009.
- [25] K.P. Burnham and D.R. Anderson. *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer, New York, USA, 2002.
- [26] California Air Resources Board. Title 13, Division 3, Chapter 1, Article 2: Approval of Motor Vehicle Pollution Control Devices (New Vehicles). *California Code of Regulations (CCR)*, 2012.

- [27] M. Castagné, Y. Bentolila, A. Hallé, F. Nicolas, and D. Sinoquet. Engine Calibration: Towards an Integrated Approach. In *Design of Experiments in Engine Development III*, number 6, Berlin, Germany, 2007.
- [28] S. Chen, S. A. Billings, and W. Luo. Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control*, 50(5):1873–1896, 1989.
- [29] S. Chen and J. Wigger. Fast orthogonal least squares algorithm for efficient subset model selection. *IEEE Transactions on Signal Processing*, 43(7):1713–1715, 1995.
- [30] F. Chmela, G. Orthaber, and W. Schuster. Die Vorausberechnung des Brennverlaufs von Dieselmotoren mit direkter Einspritzung auf der Basis des Einspritzverlaufs. *MTZ - Motor-technische Zeitschrift*, 59(7-8):484–492, 1998.
- [31] J.E. Dennis and R.B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Society for Industrial Mathematics, New Jersey, USA, 1996.
- [32] Environmental Protection Agency. Control of Air Pollution from New Motor Vehicles: Tier 2 Motor Vehicle Emissions Standards and Gasoline Sulfur Control Requirements. *Federal Register*, 65:28, 2000.
- [33] European Commission. Regulation (EC) No. 715/2007 of the European Parliament and of the Council. *Official Journal of the European Union L171*, 50:1–15, 2007.
- [34] P. Eykhoff. *System Identification: Parameter and state estimate*. Wiley-Interscience, London, UK, 1974.
- [35] L. Fahrmeir, A. Hamerle, and G. Tutz. *Multivariate statistische Verfahren*. de Gruyter, Berlin Heidelberg, Germany, 1996.
- [36] M. Farina and L. Piroddi. Simulation error minimization identification based on multi-stage prediction. *International Journal of Adaptive Control and Signal Processing*, 25(5):389–406, 2011.
- [37] Gamma Technologies Inc. GT-SUITE "Virtual Engine / Powertrain / Vehicle", "[www.gtisoft.com/img/broch/broch\\_gtsuite.pdf](http://www.gtisoft.com/img/broch/broch_gtsuite.pdf)" (31. Mai 2012). Westmont, USA.
- [38] J. Gerhardt, H. Hönninger, and H. Bischof. A new approach to functional and software structure for engine management systems-BOSCH ME7. In *SAE International Congress and Exposition*, number 980801, Detroit, USA, 1998.
- [39] P.E. Gill, W. Murray, and M.H. Wright. *Practical Optimization*. Academic Press, London, UK, 1986.
- [40] S. G. Gilmour. The Interpretation of Mallows's  $C_p$ -Statistic. *The Statistician*, 45:49–56, 1996.

- [41] K. Gschweidl, H. Pfluegl, T. Fortuna, and R. Leithgoeb. Increasing the efficiency of model-based engine applications through the use of Cameo online DoE toolbox. *ATZ worldwide eMagazines*, 17–20(7-8):274, 2001.
- [42] C. Gundlach. *Entwicklung eines ganzheitlichen Vorgehensmodells zur problemorientierten Anwendung der statischen Versuchsplanung*. Dissertation, Universität Kassel, 2004.
- [43] L. Guzzella and C.H. Onder. *Introduction to modeling and control of internal combustion engine systems*. Springer Verlag, Berlin Heidelberg, Germany, 2010.
- [44] L. Györfi, M. Kohler, A. Krzyzak, and H. Walk. *A distribution-free theory of nonparametric regression*. Springer-Verlag, New York, USA, 2002.
- [45] J. Hadler, F. Rudolph, R. Dorenkamp, H. Stehr, J. Hilzendege, and S. Kranzusch. Volkswagen's New 2.0 l TDI Engine for the Most Stringent Emission Standards. *MTZ worldwide*, 69:12–18, 2008.
- [46] M. Hafner. *Stationäre und dynamische Motoroptimierung*. Dissertation, TU Darmstadt, 2001.
- [47] M. Hafner. Effiziente Applikation der Motorsteuerung mit dynamischen Modellen (Efficient Calculation of Engine Control Maps by Means of Dynamic Models). *at-Automatisierungstechnik*, 51(5/2003):213–220, 2003.
- [48] M. Hafner, M. Schüler, O. Nelles, and R. Isermann. Fast Neural Networks for Diesel Engine Control Design. *Control Engineering Practice*, 8(11):1211–1221, 2000.
- [49] C. Hametner and S. Jakubek. New Concepts for the Identification of Dynamic Takagi-Sugeno Fuzzy Models. In *IEEE Conference on Cybernetics and Intelligent Systems*, pages 185–190, Bangkok, Thailand, 2006.
- [50] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, USA, 2001.
- [51] O. E. Herrmann, M. Krüger, and S. Pischinger. Control of Boost Pressure and EGR Rate as a Means of Emission Control in Heavy Duty Engines. *MTZ - Motortechnische Zeitschrift*, 66(11):26–29, 2005.
- [52] O.E. Herrmann. *Emissionsregelung bei Nutzfahrzeugmotoren über den Luft-und Abgaspfad*. Dissertation, RWTH Aachen, 2005.
- [53] M. Heuck, M. Feldt, M. Eichhorn, and A. Horn. Modellgestütztes AGR-Management für zukünftige Luftsysteme aufgeladener Dieselmotoren. In *AUTOREG - Steuerung und Regelung von Fahrzeugen und Motoren*, pages 65–76, Baden-Baden, Germany, 2008.
- [54] J.B. Heywood. *Internal combustion engine fundamentals*. McGraw-Hill, New York, USA, 1988.

- [55] M. Hilsch, J. Lunze, R. Nitsche, and T. Arndt. Fault-Tolerant Internal Model Control of a Diesel Engine Air Path. In *6th IFAC Symposium Advances in Automotive Control*, pages 572–577, Munich, Germany, 2010.
- [56] H. Hiroyasu, T. Kadota, and M. Arai. Development and use of a spray combustion modeling to predict diesel engine efficiency and pollutant emissions: Part 1 combustion modeling. *Bulletin of JSME*, 26(214):569–575, 1983.
- [57] H. Hiroyasu, M. Miki, J. Kamiura, S. Watanabe, and T. Hiroyasu. Multi-objective optimization of diesel engine emissions and fuel economy using genetic algorithms and phenomenological model. In *SAE Powertrain and Fluid Systems Conference and Exhibition*, number 02FFL-183, San Diego, USA, 2002.
- [58] A. Hirsch, P. Kirschbaum, and E. Winklhofer. Transientkalibrierung - Bewerten der Gemischqualität. *MTZ - Motortechnische Zeitschrift*, 69(9):708–713, 2008.
- [59] M. Hirsch and L. del Re. Sequential Identification of Engine Subsystems by Optimal Input Design. In *9th International Conference on Engines and Vehicles*, number 2009-24-0132, Capri, Naples, Italy, 2009.
- [60] A.E. Hoerl and R.W. Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67, 1970.
- [61] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.
- [62] J. Hoschek and D. Lasser. *Fundamentals of computer aided geometric design*. AK Peters, Wellesley, USA, 1993.
- [63] R. Isermann. *Digital control systems: vol. 1: fundamentals, deterministic control*. Springer-Verlag, Berlin Heidelberg, Germany, 1989.
- [64] R. Isermann. *Mechatronic Systems: Fundamentals*. Springer-Verlag, New Jersey, USA, 2005.
- [65] R. Isermann, S. Ernst (Töpfer), and O. Nelles. Identification with Dynamic Neural Networks - Architectures, Comparisons, Applications. In *11th IFAC Symposium on System Identification*, pages 997–1022, Kitakyushu, Japan, 1997.
- [66] R. Isermann, H. Kötter, and A. Schreiber. Vermessungsstrategien zur Modellbildung von Verbrennungsmotoren mit Prüfständen. In *Autoreg - Steuerung und Regelung von Fahrzeugen und Motoren*, pages 531–544, Baden-Baden, Germany, 2008.
- [67] R. Isermann and M. Münchhof. *Identification of Dynamic Systems*. Springer Verlag, Berlin Heidelberg, Germany, 2011.

- [68] J. E. Jackson. *A user's guide to principal components*. John Wiley & Sons, New York, USA, 2003.
- [69] S. Jakubek and C. Hametner. Identification of neurofuzzy models using GTLS parameter estimation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(5):1121–1133, 2009.
- [70] S. Jakubek and N. Keuth. A local neuro-fuzzy network for high-dimensional models and optimization. *Engineering Applications of Artificial Intelligence*, 19(6):705 – 717, 2006.
- [71] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, USA, 2002.
- [72] A. Juditsky, H. Hjalmarsson, A. Benveniste, B. Delyon, L. Ljung, J. Sjöberg, and Q. Zhang. Nonlinear black-box models in system identification: Mathematical foundations. *Automatica*, 31(12):1725–1750, 1995.
- [73] M. Jung. *Mean-Value Modelling and Robust Control of the Airpath of a Turbocharged Diesel Engine*. PhD thesis, University of Cambridge, 2003.
- [74] F.T. Kampelmühler, R. Paulitsch, and K. Gschweidl. Automatic ECU-Calibration - An Alternative to Conventional Methods. In *SAE International Congress and Exposition*, number 930395, Detroit, USA, 1993.
- [75] K. Keller. Kennfeldstrukturen zur Steuerung und Regelung stationärer und dynamischer Prozesse eines Dieselmotors. Master's thesis, TU Darmstadt, 2012.
- [76] J. Kennedy and R. Eberhart. Particle swarm optimization. In *IEEE International Conference on Neural Networks*, pages 1942–1948, Perth, Australia, 1995.
- [77] Y. Kidoguchi, M. Sanda, and K. Miwa. Experimental and theoretical optimization of combustion chamber and fuel distribution for the low emission direct-injection diesel engine. *Journal of engineering for gas turbines and power*, 125:351, 2003.
- [78] J. Kiefer and J. Wolfowitz. The equivalence of two extremum problems. *Canadian Journal of Mathematics*, 12:363–366, 1960.
- [79] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.
- [80] C. Knobel, K. Röpke, B. Barzantny, and S. Schaum. DoE Model driven Alternatives to Map-Based ECU software structures. In *Design of Experiments (DoE) in Engine Development II*, pages 379–388, Berlin, Germany, 2005.
- [81] B.U Köhler, W. Baumann, K. Röpke, and D. Reppel. Implementation of Statistical Models in an ECU. In *6th Conference: Design of Experiments (DoE) in Engine Development*, pages 225–234, Berlin, Germany, 2011.

- [82] M. Kohlhase. *Brennraumdruckbasiertes Motormanagement für Otto und Dieselmotoren zur Verbrauchs- und Emissionsreduktion*. Dissertation, TU Darmstadt, 2011.
- [83] M. Kohlhase, K. Pfeil, R. Zimmerschied, and R. Isermann. Cylinder-pressure-based Closed-Loop-Control of the (partial-) Homogeneous Diesel Combustion for Passenger Cars using Data-based Nonlinear Models. In *9th International Symposium on Advanced Vehicle Control*, Kobe, Japan, 2008.
- [84] P. Kotman, M. Bitzer, and A. Kugi. Flatness-Based Feedforward Control of a Diesel Engine Air System with EGR. In *6th IFAC Symposium Advances in Automotive Control*, Munich, Germany, 2010.
- [85] H. Kötter and R. Isermann. Innovative Motorvermessung - Schnelle Motorvermessung mit optimierten Testsignalen – Abschlussbericht Vorhaben Nr. 921. In *FVV Frühjahrstagung*, number Heft 853, pages 1–87, Frankfurt am Main, Germany, 2008. Forschungsvereinigung Verbrennungskraftmaschinen e.V. (FVV).
- [86] T. Kruse, S. Kurz, and T. Lang. Modern Statistical Modeling and Evolutionary Optimization Methods for the Broad Use in ECU Calibration. In *6th IFAC Symposium Advances in Automotive Control*, Munich, Germany, 2010.
- [87] T. Kruse, H. Ulmer, and U. Schulmeister. Use of Advanced Modelling and Optimization for Diesel- and Gasoline Engine Calibration. In *4th Conference: Design of Experiments (DoE) in Engine Development*, pages 91–102, Berlin, Germany, 2007.
- [88] J. Kuder and T. Kruse. Parameteroptimierung an Ottomotoren mit Direkteinspritzung. *MTZ - Motortechnische Zeitschrift*, 61(6):378–385, 2000.
- [89] R. Leithgöb, M. Bollig, M. Büchel, and F. Henzinger. Methodology for efficient calibration of model based ECU structures. In *1st International Symposium on Development Methodology*, pages 194–209, Wiesbaden, Germany, 2005.
- [90] R. Leonhard and J. Warga. Common-Rail-System von Bosch mit 2000 bar Einspritzdruck für PKW. *MTZ - Motorentechnische Zeitschrift*, 69:834–840, 2008.
- [91] I.J. Leontaritis and S.A. Billings. Input-output parametric models for non-linear systems. I: Deterministic non-linear systems. *International Journal of Control*, 41(2):303–328, 1985.
- [92] L. Ljung. *System Identification: Theory for the user*. Prentice-Hall, Inc., New Jersey, USA, 1999.
- [93] L. Ljung. Some aspects on nonlinear system identification. In *14th IFAC Symposium on System Identification*, pages 110–121, Newcastle, Australia, 2006.
- [94] L. Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.

- [95] L. Ljung and T. Glad. *Modeling of dynamic systems*. Prentice-Hall, Inc., New Jersey, USA, 1994.
- [96] C. Loader. *Local Regression and Likelihood*. Springer Verlag, New York, USA, 1999.
- [97] C. Mallows. Some comments on  $C_p$ . *Technometrics*, 15(4):661–675, 1973.
- [98] D.W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.
- [99] D.W. Marquardt. Generalized inverses, ridge regression, biased linear estimation, and non-linear estimation. *Technometrics*, 12:591–612, 1970.
- [100] E. Martini, H. Voss, S. Topfer, and R. Isermann. Effiziente Motorapplikation mit lokal linearen neuronalen Netzen. *MTZ - Motortechnische Zeitschrift*, 64(5):406–413, 2003.
- [101] G.P. Merker and C. Schwarz. *Grundlagen Verbrennungsmotoren*. Vieweg und Teubner, Wiesbaden, Germany, 2009.
- [102] A.J. Miller. *Subset Selection in Regression*. CRC Press, Florida, USA, 2002.
- [103] T. Minami, K. Takeuchi, and N. Shimazaki. Reduction of diesel engine NOx using pilot injection. In *SAE International Congress and Exposition*, number 950611, Detroit, USA, 1995.
- [104] A. Mitterer and F. Zuber-Goos. Modellgestützte Kennfeldoptimierung - Ein neuer Ansatz zur Steigerung der Effizienz in der Steuergeräteapplikation. *ATZ - Automobiltechnische Zeitschrift*, 3:188–196, 2000.
- [105] K. Mollenhauer and H. Tschoeke. *Handbook of Diesel Engines*. Springer Verlag, Berlin Heidelberg, Germany, 2010.
- [106] J. E. Moody. The Effective Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems. *Advances in Neural Information Processing Systems*, 4:847–854, 1991.
- [107] M. Mrosek and R. Isermann. Dynamic Reference Value Generation for the Control of a Diesel Engine with HP and LP-EGR. In *ASME Dynamic Systems and Control Conference*, Cambridge, USA, 2010.
- [108] M. Mrosek and R. Isermann. Model-based Estimation and Control of the Intake Gas Composition for Turbocharged Diesel Engines with High- and Low-Pressure-EGR. In *10th Stuttgart International Symposium*, Stuttgart, Germany, 2010.
- [109] M. Mrosek, S. Zahn, and I. Isermann. Parameter Estimation for Physical Based Air Path Models of Turbocharged Diesel Engines - An Experience Based Guidance. *SAE International Journal of Engines*, 2:570–583, 2010.

- [110] D. Müller. Entwurf einer nichtlinearen Druckregelung mit speziellem Störübertragungsverhalten für ein Common-Rail-Einspritzsystem. Master's thesis, TU Darmstadt, 2009.
- [111] N. Müller. *Addaptive Motorregelung beim Ottomotor unter Verwendung von Brennraumdrucksensoren*. Dissertation, TU Darmstadt, 2003.
- [112] R. H. Myers. *Classical and Modern Regression with Applications*. Duxbury Press, Boston, USA, 1990.
- [113] K. Nakakita, T. Kondoh, K. Ohsawa, T. Takahashi, and S. Watanabe. Optimization of pilot injection pattern and its effect on diesel combustion with high-pressure injection. *JSME Fluids and thermal engineering*, 37(4):966–973, 1994.
- [114] K.S. Narendra and K. Parthasarathy. Identification and control of dynamical systems using neural networks. *IEEE Transactions on Neural Networks*, 1(1):4–27, 1990.
- [115] T. Naumann. *Wissensbasierte Optimierungsstrategien für elektronische Steuergeräte an Common-Rail-Dieselmotoren*. Dissertation, TU Berlin, 2002.
- [116] J. A. Nelder and R. Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.
- [117] O. Nelles. LOLIMOT - Lokale, Lineare Modelle zur Identifikation nichtlinearer, dynamischer Systeme. *at - Automatisierungstechnik*, 45:163–174, 1997.
- [118] O. Nelles. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Springer Verlag, Berlin Heidelberg, Germany, 2001.
- [119] O. Nelles, O. Bänfer, J. Kainz, and J. Beer. Local Model Networks - The Prospective Method for Modeling in Electronic Control Units? *ATZelectronic - worldwide*, 3:36–39, 2008.
- [120] O. Nelles and A. Fink. Grid-based look-up table optimization toolbox. In *12th IFAC Symposium on System Identification*, Santa Barbara, USA, 2000.
- [121] O. Nelles, A. Fink, R. Babuska, and M. Setnes. Comparison of two construction algorithms for Takagi-Sugeno fuzzy models. *International Journal of Applied Mathematics and Computer Science*, 10(4):835–855, 2000.
- [122] O. Nelles, A. Fink, and R. Isermann. Local linear model trees (LOLIMOT) toolbox for nonlinear system identification. In *12th IFAC Symposium on System Identification*, pages 845–850, Santa Barbara, USA, 2000.
- [123] O. Nelles and R. Isermann. Basis Function Networks for Interpolation of Local Linear Models. In *35th IEEE Conference on Decision and Control*, pages 470–475, Kobe, Japan, 1996.

- [124] H.G. Nitzke, T. Rebohl, and H. Jelden. Emissionsoptimierte Abgasrückführkonzepte und dynamische Luftmassenmessung für direkteinspritzende PKW-Dieselmotoren. In *2. Symposium Steuerungssystem für den Antriebsstrang von Kraftfahrzeugen*, Berlin, Germany, 1999.
- [125] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer Verlag, New York, USA, 1999.
- [126] R. Oldenhuis. GODLIKE - A robust single-& multi-objective optimizer. MATLAB File Exchange – [www.mathworks.com/matlabcentral/fileexchange/24838](http://www.mathworks.com/matlabcentral/fileexchange/24838) (1. October 2012), August 2009.
- [127] L. Piroddi, M. Farina, and M. Lovera. Polynomial NARX Model Identification: a Wiener-Hammerstein Benchmark. In *15th IFAC Symposium on System Identification*, Saint-Malo, France, 2009.
- [128] L. Piroddi and W. Spinelli. An identification algorithm for polynomial NARX models based on simulation error minimization. *International Journal of Control*, 76(17):1767–1781, 2003.
- [129] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes: the art of scientific computing*. Cambridge University Press, New York, USA, 2007.
- [130] F. Pukelsheim. *Optimal design of experiments*. Society for Industrial Mathematics, New York, USA, 2006.
- [131] J. I. Ramos. *Internal combustion engine modeling*. Hemisphere Publishing Corporation, New York, USA, 1989.
- [132] I. Rechenberg. *Evolutionsstrategie*. Friedrich Frommann Verlag, Stuttgart, Germany, 1973.
- [133] P. Renninger and M. Aleksandrov. Rapid Hull Determination: a new method to determine the design space for model based approaches. In *Design of Experiments (DOE) in Engine Development II*, pages 14 – 29, Berlin, Germany, 2005.
- [134] P. Renninger, M. Weirich, K. Pfeil, and R. Isermann. Optimierungsstrategien für den transienten Betrieb eines Dieselmotors innerhalb der Rauchbegrenzung. In *ATZ/MTZ-Konferenz – Motorenentwicklung auf dynamischen Prüfständen*, pages 1–12, Wiesbaden, Germany, 2006.
- [135] D.E. Rivera, M. Morari, and S. Skogestad. Internal model control: PID controller design. *Industrial & engineering chemistry process design and development*, 25(1):252–265, 1986.
- [136] Robert-Bosch-GmbH. *Diesel-engine management*. Wiley, Chichester, UK, 2006.
- [137] K. Röpke. *Design of Experiments (DOE) in der Motorenentwicklung: Praktische Anwendungen in der Automobil-und Zuliefererindustrie*. expert-Verlag GmbH, Renningen, Germany, 2003.

- [138] K. Röpke, M. Knaak, B. Barzantny, U. Genc, and U. Plewnia. Future methods for a model-based engine development and calibration process. In *1. International Symposium on Development Methodology*, pages 90–101, Wiesbaden, Germany, 2005.
- [139] K. Röpke, M. Knaak, A. Neßler, and S. Schaum. Rapid Measurement - Basic Combustion Engine Calibration In One Day? *MTZ - worldwide*, 04:16–19, 2007.
- [140] J. Rückert. *Modellgestützte Regelung von Ladedruck und Abgasrückführrate beim Dieselmotor*. Dissertation, RWTH Aachen, 2005.
- [141] J. Ruckert, B. Kinoo, M. Kruger, A. Schlosser, H. Rake, and S. Pischinger. Simultaneous control of boost pressure and exhaust gas recirculation in a passenger car diesel engine. *MTZ - worldwide*, 62(11):32–35, 2001.
- [142] T.J. Santner, B.J. Williams, and W. Notz. *The design and analysis of computer experiments*. Springer Verlag, New York, USA, 2003.
- [143] A. Schlosser. *Modellbildung und Simulation zur Ladedruck- und Abgasrückführregelung an einem Dieselmotor*. Dissertation, RWTH Aachen, 2000.
- [144] A. Schlosser, C. Schönfelder, M. Hendriks, S Pischinger, and T. Sentis. Automated ECU-Calibration - Example: Torque Structure of Gasoline Engine. In *5th Design of Experiments (DoE) in Engine Development System*, number 14, Berlin, Germany, 2009.
- [145] M. Schmitt. *Untersuchung zur Realisierung mehrdimensionaler lernfähiger Kennfelder in Großserien-Steuergeräten*. Dissertation, TU-Darmstadt, 1995.
- [146] G. Schopp, T. Burkhardt, J. Dingl, R. Schwarz, and C. Eisath. Funktionsentwicklung und Kalibration für aufgeladene Motoren - Modellbasiert vom Konzept bis zur Serie. In R. Isermann, editor, *Elektronisches Management motorischer Fahrzeugantriebe*, chapter 9, pages 214–232. Springer Verlag, Berlin Heidelberg, Germany, 2010.
- [147] D. Schöppe, S. Zulch, M. Hardy, D. Geurts, R.W. Jorach, and N. Baker. Common-Rail-Einspritzung mit Direct-Acting-Injektor von Delphi. *MTZ - Motortechnische Zeitschrift*, 69(10):842, 2008.
- [148] A. Schreiber, R. Isermann, and M.-S. Vogels. Dynamische Vermessung und Modellbildung hochdimensionaler Zusammenhänge im Verbrennungsmotor mit dynamischen Motorenprüfständen. In *2. Internationales Symposium für Entwicklungsmethodik*, pages 26–39, Wiesbaden, Germany, 2007.
- [149] M. Schüler. *Stationäre Optimierung der Motorsteuerung von PKW-Dieselmotoren mit Abgasturbolader durch Einsatz schneller neuronaler Netze*. Dissertation, TU Darmstadt, 2001.
- [150] M. Schüler, M. Hafner, and R. Isermann. Einsatz schneller neuronaler Netze zur modellbasierten Optimierung von Verbrennungsmotoren. *MTZ Motortechnische Zeitschrift*, 61(10), 2000.

- [151] M. Schüler, M. Hafner, and R. Isermann. Model-based optimization of IC engines by means of fast neural networks, Part 1: Modelling the engine- and emission behavior. *MTZ worldwide*, 61(10/11):10–11, 2000.
- [152] A. Schwarte, L. Hack, I. Isermann, H.G. Nitzke, J. Jeschke, and J Piewek. Automatisierte Applikation von Motorsteuergeräten mit kontinuierlicher Motorvermessung. In *Autoreg - Steuerung und Regelung von Kraftfahrzeugen und Verbrennungsmotoren*, Wiesloch, Germany, 2004.
- [153] H.P. Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Birkhäuser Verlag, Basel-Stuttgart, 1977.
- [154] G. A. F. Seber. *Linear Regression Analysis*. John Wiley & Sons Inc, New Jersey, USA, 1977.
- [155] P.K. Senecal, D.T. Montgomery, and R.D. Reitz. A methodology for engine design using multi-dimensional modelling and genetic algorithms with validation through experiments. *International Journal of Engine Research*, 1(3):229–248, 2000.
- [156] Y. Shinohara, K. Takeuchi, O.E. Herrmann, and HJ LAUMEN. Common-Rail-Einspritzsystem mit 3000 Bar. *MTZ - Motortechnische Zeitschrift*, 72(1):10–15, 2011.
- [157] J. Sjöberg, Q. Zhang, L. Ljung, A. Benveniste, B. Delyon, P.Y. Glorennec, H. Hjalmarsson, and A. Juditsky. Nonlinear black-box modeling in system identification: a unified overview. *Automatica*, 31(12):1691–1724, 1995.
- [158] R. Steinbrink, D. Überschär, J. Kopp, and G. Eifler. Verbesserung des Kraftstoffverbrauchs von Dieselmotoren durch den Einsatz eines rechnergestuetzten Kennfeldoptimierungssystems Teil 2. *Motortechnische Zeitschrift*, 63(11&12):1052–1057, 2002.
- [159] G. Stiesch. Modellansätze für die Simulation von Gemischbildung und Verbrennung. In R. Isermann, editor, *Elektronisches Management motorischer Fahrzeugantriebe*, chapter 4, pages 88–102. Springer Verlag, Berlin Heidelberg, Germany, 2010.
- [160] E. Stoelting, J. Seebode, R. Gratzke, and K. Behnk. Regeln und Steuern: Emissionsgeführtes Motormanagement für Nutzfahrzeuganwendungen. *MTZ - Motortechnische Zeitschrift*, 69(12):30–35, 2008.
- [161] P. Stoica and Y. Selen. Model-Order Selection: A review of information criterion rules. *Signal Processing Magazine, IEEE*, 21(4):36–47, 2004.
- [162] H. Stuhler, K. Gschweidl, T. Kruse, A. Stuber, W. Piock, and H. Pfluegl. Automated model-based GDI engine calibration adaptive online DoE approach. In *SAE World Congress and Exhibition*, number 2002-01-0708, Detroit, USA, 2002.
- [163] H Stuhler and T. Kruse. Parametrierung von Motorsteuerungen. In R. Isermann, editor, *Modellgestützte Steuerung, Regelung und Diagnose von Verbrennungsmotoren*, chapter 3, pages 23–38. Springer Verlag, Berlin Heidelberg, Germany, 2003.

- [164] A. N. Tikhonov and V. I. A. Arsenin. *Solutions of ill-posed problems*. Winston & Sons, Washington DC, USA, 1977.
- [165] S. Töpfer. Realisation of Hierarchical Look-up Tables on Low-Cost Hard-ware. In *12th IFAC Symposium on System Identification*, Santa Barbara, USA, 2000.
- [166] S. Töpfer. *Hierarchische neuronale Modelle für die Identifikation nichtlinearer Systeme*. Dissertation, TU Darmstadt, 2002.
- [167] S. Töpfer, E. Martini, and C. Kunde. Stationäre Motorvermessung mit neuronalen Netzen. In R. Isermann, editor, *Modellgestützte Steuerung, Regelung und Diagnose von Verbrennungsmotoren*, chapter 9, pages 131–152. Springer Verlag, Berlin Heidelberg, Germany, 2003.
- [168] S. Töpfer and O. Nelles. Polynommodelle, Kennfelder und neuronale Netze. In R. Isermann, editor, *Modellgestützte Steuerung, Regelung und Diagnose von Verbrennungsmotoren*, chapter 7, pages 103–119. Springer Verlag, Berlin Heidelberg, Germany, 2003.
- [169] F. Tschanz, C. Barro, A. Amstutz, L. Guzzella, and K. Boulouchos. Regelungsorientierte Modellbildung der Russemissionen und modellbasierte Regelung der Russemissionen im Motorbetrieb - Abschlussbericht Vorhaben Nr. 986. In *FVV Frühjahrstagung*, volume Heft R558, Bad Neuenahr, Germany, 2012.
- [170] F. Tschanz, C. Barro, A. Amstutz, P. Obrecht, K. Boulouchos, and Guzzella L. On-line simulation and feedback control of pollutant emissions from Diesel engines. In *11th Stuttgart International Symposium*, pages 469–486, Stuttgart, Germany, 2011.
- [171] T. Ullrich. *Untersuchungen zur effizienten interpolierenden Speicherung von nichtlinearen Prozeßmodellen und Vorsteuerstrategien*. Dissertation, TU Darmstadt, 1998.
- [172] S. Unland, H. Stuhler, and A. Stuber. Improved calibration process for the physically based engine management control system ME7 (german). *MTZ - Motortechnische Zeitschrift*, 59(11):744–751, 1998.
- [173] S. Unland, H. Stuhler, and A. Stuber. Neue effiziente Applikationsverfahren für die physikalisch basierte Motorsteuerung ME7. *MTZ - Motortechnische Zeitschrift*, 59(11):744–751, 1998.
- [174] M. van Nieuwstadt, P. E. Moraal, I. Kolmanovsky, A. Stefanopoulou, P. Wood, and M. Criddle. Decentralized and multivariable designs for EGR-VGT control of a diesel engine. In *IFAC Workshop on Advances in Automotive Control*, Mohican State Park, USA, 1998.
- [175] M. J. van Nieuwstadt, I. V. Kolmanovsky, P. E. Moraal, A. Stefanopoulou, and M. Jankovic. EGR-VGT control schemes: Experimental comparison for a high-speed diesel engine. *Control Systems Magazine, IEEE*, 20(3):63–79, 2000.

- [176] V. Vapnik, S. E. Golowich, and A. Smola. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. In *Advances in Neural Information Processing Systems 9*, pages 281–287. MIT Press, 1996.
- [177] VDI-Gesellschaft Verfahrenstechnik und Ingenieurwesen. *VDI-Wärmeatlas*. Springer, Berlin Heidelberg, Germany, 2006.
- [178] I. I. Vibe. *Brennverlauf und Kreisproze von Verbrennungsmotoren*. VEB Verlag Technik, Berlin, Germany, 1970.
- [179] M. Vogt. *Support Vector Machines for Identification and Classification Problems in Control Engineering*. Dissertation, TU Darmstadt, 2008.
- [180] M. Vogt, N. Müller, and R. Isermann. On-line adaptation of grid-based look-up tables using a fast linear regression technique. *Journal of dynamic systems, measurement, and control*, 126(4):732–739, 2004.
- [181] K. von Pfeil. *Ladedruck- und Luftmassenregelung von aufgeladenen Dieselmotoren mit lokal linearen Modellen und Optimierung des dynamischen Emissionsverhaltens im Rauchbetrieb*. Dissertation, TU Darmstadt, 2011.
- [182] X. Wei, L. del Re, and P. Langthaler. Lpv dynamical models of diesel engine nox emission. In *IFAC Symposium Advances in Automotive Control*, pages 245–250, Salerno, Italy, 2004.
- [183] E. Winklhofer, M. Neubauer, A. Hirsch, and H. Philipp. Cylinder and Cycle Resolved Particle Formation Evaluation to Support GDI Engine Development for Euro 6 Targets. In *10th International Conference on Engines and Vehicles*, number 2011-24-0206, Capri, Naples, Italy, 2011.
- [184] X. Yin, F. Campean, A. Wood, B. Seale, and M. Goodman. Automatic Optimisation Strategies for Diesel Engine Calibration. In *5th Conference: Design of Experiments (DoE) in Engine Development*, number 16, Berlin, Germany, 2009.
- [185] S. Zahn. Mittelwert- und Arbeitstaktsynchrone Simulation von Dieselmotoren. In R. Isermann, editor, *Elektronisches Management motorischer Fahrzeugantriebe*, chapter 6, pages 130–166. Springer Verlag, Berlin Heidelberg, Germany, 2010.
- [186] R. Zimmerschied. *Entwurf von Anregungssignalen für die Identifikation nichtlinearer dynamischer Prozesse*. Diploma thesis, TU Darmstadt, 2002.
- [187] R. Zimmerschied. *Identifikation nichtlinearer Prozesse mit dynamischen lokal-affinen Modellen*. Dissertation, TU Darmstadt, 2008.
- [188] R. Zimmerschied and R. Isermann. Regularisierungsverfahren für die Identifikation mittels lokal-affiner Modelle. *at-Automatisierungstechnik*, 56:339–349, 2008.

- [189] R. Zimmerschied, M. Weber, and R. Isermann. Static and Dynamic Measurements of Combustion Engines for Optimization of Control Mappings - A Brief Survey. *at-Automatisierungstechnik*, 53(2/2005):87–94, 2005.

---

## Publications by the author

- [190] R. Isermann, M. Kowalczyk, F. Kunkel, A. Schreiber, and H. Sequenz. Possibilities for target-oriented application of dynamic models in engine calibration. In *4th International Symposium on Development Methodology*, pages 182–195, Wiesbaden, Germany, 2011.
- [191] M. Kowalczyk and H. Sequenz. Stationäre und dynamische Motorvermessung mit verschiedenen Methoden und Modellen. In *Haus der Technik – Elektronisches Management von Fahrzeugantrieben*, number 7, Frankfurt am Main, Germany, 2012.
- [192] H. Kötter and H. Sequenz. Stationäre Motorvermessung mit verschiedenen Methoden und Modellen. In R. Isermann, editor, *Elektronisches Management motorischer Fahrzeugantriebe*, chapter 6, pages 130–166. Springer Verlag, Berlin Heidelberg, Germany, 2010.
- [193] H. Kötter, H. Sequenz, and R. Isermann. Ermittlung der Güte experimentell gewonnener Verbrennungsmotor-Modellen. In *2nd International Symposium on Development Methodology*, pages 94–113, Wiesbaden, Germany, 2007.
- [194] M. Mrosek, H. Sequenz, and R. Isermann. Control Oriented NO<sub>x</sub> and Soot Models for Diesel Engines. In *6th IFAC Symposium Advances in Automotive Control*, Munich, Germany, 2010.
- [195] M. Mrosek, H. Sequenz, and R. Isermann. Identification of Emission Measurement Dynamics for Diesel Engines. In *18th IFAC World Congress*, Milano, Italy, 2011.
- [196] M. Mrosek, H. Sequenz, and R. Isermann. Transiente Luftpfadregelung für Dieselmotoren - Ein modellbasierter Ansatz für optimierte Sollwerte. In *Autoreg - Steuerung und Regelung von Fahrzeugen und Motoren*, Baden-Baden, Germany, 2011.
- [197] M. Mrosek, H. Sequenz, and R. Isermann. Transiente Luftpfadregelung für Dieselmotoren - Ein modellbasierter Ansatz für optimierte Sollwerte. In *2. Automobiltechnisches Kolloquium München*, Munich, Germany, 2011.
- [198] H. Sequenz. Automatische Struktursuche für dynamische Systeme mit Regressionsmodellen. Diploma thesis, TU Darmstadt, 2007.
- [199] H. Sequenz and R. Isermann. Calibration of Diesel Engine Control Systems with Regard to Series Variations and Model-Uncertainties. In *Autoreg - Steuerung und Regelung von Fahrzeugen und Motoren*, Baden-Baden, Germany, 2011.
- [200] H. Sequenz and R. Isermann. Emission Model Structures for the Implementation on Engine Control Units. In *18th IFAC World Congress*, Milano, Italy, 2011.
- [201] H. Sequenz and R. Isermann. Local Adaptive Polynomials for Identification of Stationary and Dynamic Systems (german). *at-Automatisierungstechnik*, accepted.

- 
- [202] H. Sequenz, K. Keller, and R. Isermann. About the Identification of Higher Dimensional Look-Up Tables for Combustion Engines (german). *at-Automatisierungstechnik*, 60, 2012.
- [203] H. Sequenz, M. Mrosek, and R. Isermann. A Global-Local Emission-Model for NO<sub>x</sub> and Soot Emissions of Turbocharged CR-Diesel Engines. In *ASME Dynamic Systems and Control Conference*, Cambridge, USA, 2010.
- [204] H. Sequenz, M. Mrosek, and R. Isermann. Stationary Global-Local Emission-Models of a CR-Diesel Engine with Adaptive Regressor-Selection for Measurements of Airpath and Combustion. In *6th IFAC Symposium Advances in Automotive Control*, Munich, Germany, 2010.
- [205] H. Sequenz, M. Mrosek, and R. Isermann. Model Based Smoke Limitation for the Control of Turbocharged CR-Diesel Engines. In *11th Stuttgart International Symposium*, Stuttgart, Germany, 2011.
- [206] H. Sequenz, M. Mrosek, S. Zydek, and R. Isermann. Model Based Optimization of a Step in Acceleration for a CR-Diesel Engine. In *18th IFAC World Congress*, Milano, Italy, 2011.
- [207] H. Sequenz, A. Schreiber, and R. Isermann. Identification of Nonlinear Static Processes with Local Polynomial Regression and Subset Selection. In *15th IFAC Symposium on System Identification*, Saint-Malo, France, 2009.