



Doris Schadler, BSc

**Modellbasierte Methoden  
zur Fehlerdiagnose  
an Motorprüfständen**

**MASTERARBEIT**

zur Erlangung des akademischen Grades

Diplom-Ingenieurin

Masterstudium Technische Mathematik: Operations Research und Statistik

eingereicht an der

**Technischen Universität Graz**

Betreuer

Univ.-Prof. Dipl.-Ing. Dr.techn. Ernst Stadlober

Institut für Statistik



## **EIDESSTATTLICHE ERKLÄRUNG**

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Das in TUGRAZonline hochgeladene Textdokument ist mit der vorliegenden Masterarbeit identisch.

---

Datum

---

Unterschrift



## Kurzfassung

Die Anforderungen an die Messdatenqualität in der Motorenentwicklung wachsen durch den steigenden Kosten- und Zeitdruck immer mehr. Daher ist eine Überprüfung der Messdatenqualität unmittelbar am Prüfstand mit einem Diagnoseverfahren unumgänglich. Im Rahmen dieser Arbeit wird dies mittels statistischer Modelle auf Basis der Multiplen Linearen Regression und unterschiedlicher Prognosegütemaße versucht. Bei der Findung der Modelle können durch das Modellauswahlverfahren der Schrittweisen Regression und Transformationen erklärender Größen gute Beschreibungen der Zielgrößen erreicht werden. Das Diagnoseverfahren basiert auf diesen Modellberechnungen und vergleicht die Prognosewerte mit den Messdaten anhand von drei Kriterien. Liegen die relativen Prognosefehler außerhalb der durch die Kriterien definierten Bandbreiten, dann werden die dazugehörigen Messpunkte als fehlerhaft eingestuft. Bei der Evaluierung mittels Datensätzen mit simulierten Fehlern liefert das Verfahren durchwegs zufriedenstellende Ergebnisse. Lediglich bei Messgrößen mit einem hohen Variationskoeffizienten werden systematische Fehler durch den zufälligen Fehler überlagert und somit vom Verfahren nicht als solche erfasst. Die Masterarbeit entstand in Kooperation mit dem Large Engines Competence Center (LEC) unter der Betreuung von DI Dr. Gerhard Pirker und DI Michael Wohlthan.

## Abstract

The requirements for the quality of measurement data in engine development are growing due to the increasing cost and time pressures constantly. Therefore, a review of data quality with a diagnosis method immediately on the test bench is essential. The method presented in this paper is based on the theory of multiple linear regression and different forecast accuracy measures. The choice of predictive variables with the automatic procedure called stepwise regression and transformations of certain variables lead to models with an appropriate goodness of fit. The fault diagnostic is based on these models and compares the predicted values with measured data based on three criteria. If the relative forecast errors realise outside the ranges defined by the criteria, the datapoints according to them are assumed to be defective. The evaluation of the method on the basis of data sets with simulated faults yields satisfying results. Only at measurements with a high coefficient of variation random errors overlay systematic errors, so that they don't become detected by the system. This master thesis has been written in cooperation with the Large Engines Competence Center (LEC) under the supervision of DI Dr. Gerhard Pirker and DI Michael Wohlthan.



# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b>	<b>x</b>
<b>Tabellenverzeichnis</b>	<b>xi</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Motorprüfstand . . . . .	1
1.2 Aufgabenstellung . . . . .	4
<b>2 Datenmaterial</b>	<b>7</b>
2.1 Datengrundlage . . . . .	7
2.2 Messgrößen . . . . .	8
2.3 Motorkonfiguration . . . . .	10
2.4 Datenaufbereitung . . . . .	11
<b>3 Explorative Datenanalyse</b>	<b>15</b>
3.1 Univariate Datenanalyse . . . . .	15
3.2 Bivariate Datenanalyse . . . . .	20
<b>4 Modelle</b>	<b>25</b>
4.1 Multiple lineare Regressionsmodelle . . . . .	25
4.2 Kleinste-Quadrate-Schätzung . . . . .	27
4.3 Modellselektion . . . . .	30
4.4 Indikatorvariablen . . . . .	31
4.5 Modellvalidierung . . . . .	33
4.6 Resultate der praktischen Anwendung . . . . .	34
4.6.1 Modellfindung . . . . .	34
4.6.2 Sonderfall AbgNOx . . . . .	39
4.6.3 Konfigurationenanalyse . . . . .	40
4.6.4 Finale Modelle . . . . .	42
4.6.5 Evaluierung . . . . .	44

<b>5</b>	<b>Messdatendiagnose</b>	<b>47</b>
5.1	Prognose und Prognosefehler . . . . .	47
5.2	Prognosegütemaße . . . . .	49
5.3	Diagnoseverfahren . . . . .	51
5.4	Anmerkungen zur praktischen Umsetzung . . . . .	57
5.5	Vergleich der Kriterien . . . . .	59
<b>6</b>	<b>Evaluierung</b>	<b>63</b>
6.1	Alternative Methode . . . . .	63
6.2	Kennzahlen zur Bewertung . . . . .	65
6.3	Prinzip der Evaluierung . . . . .	67
6.4	Ergebnisse . . . . .	68
6.4.1	Performance der entwickelten Methode . . . . .	68
6.4.2	Vergleich der beiden Methoden . . . . .	72
<b>7</b>	<b>Resümee und Ausblick</b>	<b>75</b>
	<b>Literatur</b>	<b>77</b>

# Abbildungsverzeichnis

1.1	Schematischer Aufbau eines Motorprüfstandes . . . . .	2
1.2	Einzyylinder-Forschungsmotor an der LEC GmbH . . . . .	3
3.1	Histogramme aller Messgrößen, $n = 7945$ . . . . .	18
3.2	Boxplots ausgewählter Messgrößen, $n = 7945$ . . . . .	19
3.3	Korrelationsmatrix aller Messgrößen . . . . .	21
3.4	Scatterplotmatrix der Messgrößen <b>mLuft</b> , <b>mGas</b> , <b>IMEP</b> , <b>p2</b> und <b>M</b> . . . . .	22
3.5	Scatterplotmatrix der Abgaskomponenten . . . . .	23
4.1	Schemahafte Darstellung der Modellfindung . . . . .	35
4.2	Analyse von Scatterplots zur Modellbildung für die Zielgröße <b>AbgNOx</b> . . . . .	39
4.3	Scatterplots zur Konfigurationenanalyse von <b>mLuft</b> . . . . .	41
4.4	Scatterplots zur Konfigurationenanalyse von <b>mGas</b> . . . . .	41
4.5	Scatterplots zur Konfigurationenanalyse von <b>IMEP</b> . . . . .	41
4.6	Scatterplots zur Konfigurationenanalyse von <b>AbgCO</b> . . . . .	42
4.7	Q-Q-Plots der standardisierten Residuen . . . . .	44
4.8	Scatter-Plots: Stand. Residuen gegen Vorhersagewerte . . . . .	45
5.1	Kriterium 1 - Großer Vertrauensbereich . . . . .	52
5.2	Kriterium 1 - Kleiner Vertrauensbereich . . . . .	53
5.3	Unterschied im Wertebereich der relativen Fehler . . . . .	54
5.4	Kriterium 2 . . . . .	56
5.5	Kriterium 3 . . . . .	57
5.6	Notwendigkeit von Kriterium 1 (Großer Vertrauensbereich) . . . . .	60
5.7	Vorteil von Kriterium 2 . . . . .	61
5.8	Vorteil von Kriterium 3 . . . . .	62
6.1	Beispielhafte Aufprägungen der beiden Fehlerarten . . . . .	68
6.2	Auswertung der Kennzahlen zum Sprung je Messgröße . . . . .	69
6.3	Auswertung der Kennzahlen zum Drift je Messgröße . . . . .	70
6.4	Scatterplot: Kennzahlen bzgl. der Fehlerisolation gegen Variationskoeffizienten der Zielgrößen . . . . .	71
6.5	Auswertung der Kennzahlen zum Sprung der beiden Varianten . . . . .	72

6.6 Auswertung der Kennzahlen zum Drift der beiden Varianten . . . . . 73

# Tabellenverzeichnis

2.1	Beschreibung der analysierten Messgrößen . . . . .	9
2.2	Messreihen aufgespalten nach den Bauteilen Kolben und Nockenwelle .	12
3.1	Deskriptive Kenngrößen aller Messgrößen, $n = 7945$ . . . . .	17
4.1	Beispielhafte Modellfindung anhand der Zielgröße <b>AbgCO2</b> . . . . .	38
4.2	Analyse der Modelle der 9 Zielgrößen, $n = 7945$ . . . . .	43
5.1	Prognosegütemaße . . . . .	50
5.2	Übersicht über die Kriterien und deren Parameter . . . . .	59
6.1	Kategorien eines klassifizierten Messpunktes . . . . .	65
6.2	Beispielhafte Fehleraufprägung für einen Sprung . . . . .	67
6.3	Beispielhafte Fehleraufprägung für einen schleichenden Fehler . . . . .	68



# 1 Einleitung

Zu Beginn dieser Arbeit wird die zugrundeliegende Thematik vorgestellt. Dabei wird der Zweck und Aufbau des Motorprüfstandes erläutert und auf den in der Arbeit verwendeten Spezialfall eingegangen. Dies sowie auch etwaige Messtechniken und Messprinzipien können genauer in WIMMER (2014) nachgelesen werden. Die grundsätzliche Aufgabenstellung dieser Arbeit ist eine Folge der immer höher werdenden Ansprüche an solche Motorprüfstände. So wird die Entwicklung einer Messdatendiagnose mit der steigenden Anforderung an die Messdatenqualität durchaus immer notwendiger. Dabei wird kurz erläutert, auf welche Art und Weise diese Problematik im Rahmen der vorliegenden Arbeit behandelt wird.

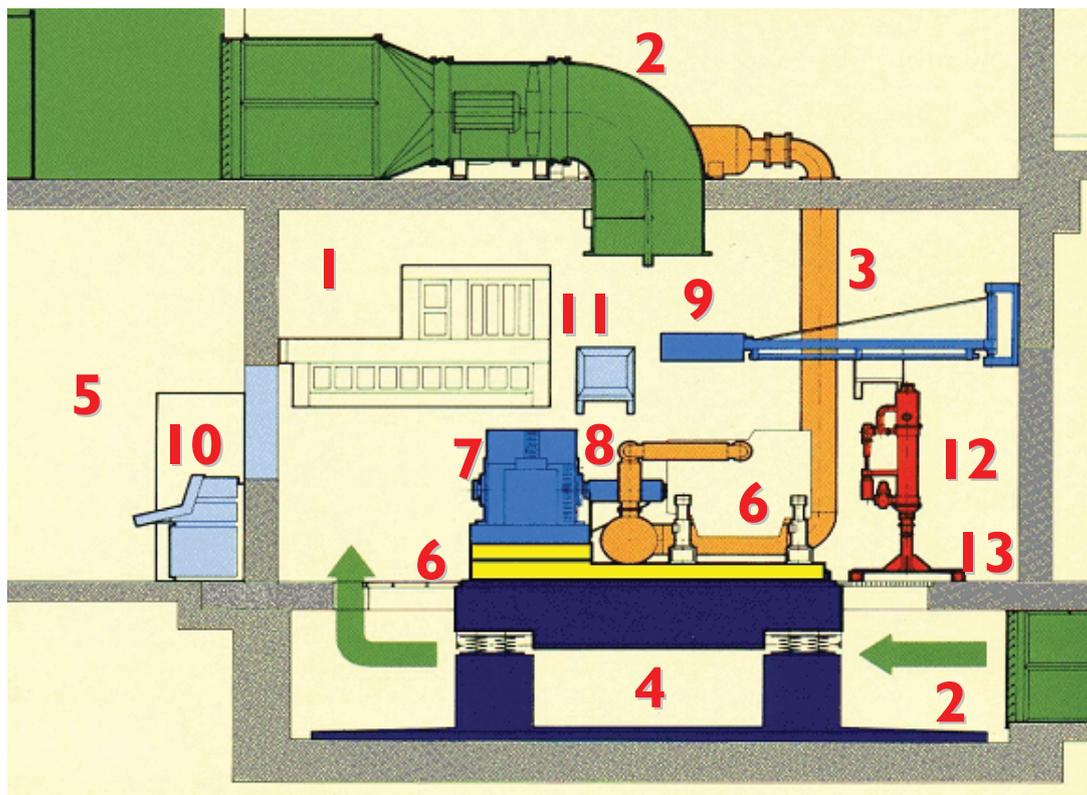
## 1.1 Motorprüfstand

In der Fahrzeug- und Motorenmesstechnik existiert eine Vielzahl an Prüfständen. Ein Prüfstand ist dabei immer für den entsprechenden Einsatzzweck konzipiert. Für die Entwicklung eines Gesamtfahrzeuges bis zur Serienreife werden die Einzelkomponenten auf speziellen Prüfständen getestet und optimiert. Dazu zählen unter anderem Untersuchungen der Verbrennungskraftmaschine am Motorprüfstand, Optimierung von Aggregaten an Aggregatprüfständen und die Verbesserung von Antriebsstrangkomponenten. Auf sogenannten Rollenprüfständen kann das Zusammenspiel aller Komponenten untersucht werden. Auch Spezialanwendungen wie etwa Fahrzeug-Windkanäle kommen zum Einsatz.

Ein Prüfstand besteht im Allgemeinen aus einem Prüfraum (Prüfzelle) und üblicherweise einem mit einer Zugangstür verbundenen Beobachtungsraum. Im Prüfraum ist je nach Prüfstandstyp der zu untersuchende Versuchsträger (z.B. Motor, Getriebe, Aggregate, Gesamtfahrzeug,...) auf speziellen Vorrichtungen und Trägerkomponenten zusammen mit der jeweils erforderlichen messtechnischen Einrichtung aufgebaut. Vom Beobachtungsraum aus erfolgt die Prüfstandbedienung und -überwachung, unter anderem über ein großflächiges Beobachtungsfenster. Außerdem beinhaltet er die Systemsteuerung und Automatisierungstechnik.

## 1 Einleitung

Bei einem Motorprüfstand befindet sich im Prüfstandsraum der Verbrennungsmotor. Dieser ist über eine Verbindungswelle mit einer Belastungsmaschine (Leistungsbremse) verbunden. Genauer gesagt, sind Motor und Bremse über eine Kupplung verbunden, damit die Spitzen der Motormassenkräfte gedämpft werden. Das gesamte System (Motor-Welle-Bremse) ist auf einem schwingungsfähigen Fundament aufgebaut, um es schwingentechnisch vom Gebäude, in dem der Prüfstand untergebracht ist, zu entkoppeln.



Legende:

1	Prüfraum	11	Motorspezifische Messtechnik; z.B.: Kraftstoffwaage Drehkolbengaszähler Blow-By- Meter Ölverbrauchsmessgerät
2	Be- und Entlüftung	12	Konditioniereinrichtungen; z.B.: Ölkonditionierung Kühlmittelkonditionierung Kraftstoffkonditionierung
3	Auspuffanlage		
4	Schwingfundament		
5	Beobachtungsraum	13	Anschluss für Sekundärkühlmittel (Hauswasser)
6	Versuchsträger (Motor)		
7	Leistungsbremse		
8	Wellenkupplung		
9	Messgalgen		
10	Bedienfeld mit: Prüfstandsautomatisierung Abgasanalyse Kalibriergasentnahme		

Abbildung 1.1: Schematischer Aufbau eines Motorprüfstandes  
(Quelle: WIMMER (2014), Kapitel 4: Seite 24)

Wie an der Abbildung des Motorprüfstandes erkennbar, gibt es an diesem eine Reihe von Konditioniereinrichtungen. Diese sind notwendig um für die Bewertung von konstruktiven Veränderungen am Motor oder Variationen der Betriebsmedien konstante Rahmenbedingungen schaffen zu können. Es können damit etwa Temperatur und Durchflussmenge des Kraftstoffes, der Ansaugluft, des Motoröles oder der Kühlflüssigkeit gemessen bzw. geregelt werden. Des Weiteren sind Einrichtungen für die motorspezifische Messtechnik angebracht und in die Auspuffanlage spezielle messtechnische Komponenten aufgenommen. Zur Anbindung der applizierten Sensoren werden Kabelverbindungen von der Prüfstandsautomatisierung bis hin zum Motor über einen Kabelgalgen geführt.

Man unterscheidet bei Motorprüfständen zwischen stationären, dynamischen Motorprüfständen und weiters Reibleistungs- und Akustikprüfständen. Die vorliegende Arbeit befasst sich mit Messdaten eines stationären Motorprüfstandes für Einzylinderforschungsmotoren. Solche Motoren dienen der Grundlagenforschung. Es handelt sich meist um Abbildungen von Mehrzylinder-Serienmotoren, von denen u.a. Zylinderkopf, Zylinderlaufbuchse und Kolben übernommen werden, um die damit erhaltenen Messergebnisse auf den Mehrzylindermotor zu übertragen. Außerdem sind bei Einzylindermotoren Umbauten und Variationen einfacher und kostengünstiger umzusetzen. Die Abbildung zeigt ein Beispiel eines Einzylinder-Forschungsmotors an der LEC GmbH (Laborgebäude Inffeldgasse 21).



Abbildung 1.2: Einzylinder-Forschungsmotor an der LEC GmbH  
(Quelle: WIMMER (2014), Kapitel 4: Seite 7)

Derartige Prüfstände kommen vor allem bei Großgas- und Großdieselmotoren zum Einsatz, etwa für die Entwicklung von Brennverfahren, Schmieröl- und Kraftstoffuntersuchungen und andere Bauteiloptimierungen. Die Anwendungsgebiete solcher Motoren sind Schiffs- und Lokomotivantriebe, aber auch Kraft- und Wärmekopplungsanlagen sowie Generatorantriebe in der Energieerzeugung.

## 1.2 Aufgabenstellung

Wie im vorhergehenden Abschnitt beschrieben, gibt es an einem Motorprüfstand eine Reihe von überwachten Sensoren bzw. Messgrößen. Der Wunsch des Anwenders besteht natürlich darin, dass sämtliche Messungen so präzise wie möglich geschehen und etwaige Fehler, Störungen oder gar Ausfälle frühzeitig erkannt werden. Dies wird vor allem dadurch motiviert, dass der Betrieb und auch ein etwaiger Umbau des Prüfstandes mit Kosten verbunden sind, die so gering wie möglich gehalten werden sollten. Die grundlegende Aufgabenstellung liegt also in der Entwicklung eines Diagnoseverfahrens, das dies ermöglicht.

Bevor jedoch ein Diagnoseverfahren an sich aufgebaut wird, muss festgelegt werden, was überhaupt unter einem Fehler zu verstehen ist. Grundsätzlich lassen sich dabei die Fehler nach WIMMER (2014) kategorisieren in grobe, systematische und zufällige Fehler. Grobe Fehler werden dabei oft vom Beobachter aus Versehen oder Unachtsamkeit verursacht. Dabei weichen die betroffenen Messwerte sehr stark von den anderen ab. Der zufällige Fehler beschreibt unbeeinflussbare Veränderungen während der Messung. Betrag und Vorzeichen sollten dabei in völliger Zufälligkeit abwechseln. Abgezielt wird in dieser Arbeit auf Detektion des systematischen Fehlers, welcher durch unvollkommene Messgeräte, Messverfahren oder Umwelteinflüsse entsteht.

Je nach ihren Eigenschaften können solche systematische Fehler noch weiter kategorisiert werden. So kann man etwa zwischen *schleichenden*, *abrupten* und *sporadischen Fehlern* unterscheiden, die laut WOLFF (2010) wie folgt definiert sind:

- Ein *schleichender Fehler* (auch Drift genannt) ist ein Fehler, dessen Ursache und Auswirkungen zunächst klein sind, die sich mit der Zeit jedoch immer mehr verstärken.
- Ein *abrupter Fehler* ist ein plötzlich auftretender Fehler, der zu einer schlagartigen Änderung des Systemverhaltens führt.

- Ein *sporadischer* Fehler ist ein Fehler, der wiederholt auftritt und wieder verschwindet.

Das Ergebnis dieser Arbeit soll ein Verfahren sein, womit derartige systematische Fehler rechtzeitig und zuverlässig erkannt werden. Das Verfahren, das im Rahmen dieser Arbeit gestaltet und entwickelt wird, beruht auf der Basis statistischer Modelle und Fehlermaße. Das bedeutet, dass zuerst adäquate (lineare) Modelle für die einzelnen Zielgrößen gefunden werden müssen, um diese dann im nächsten Schritt für die Fehlerdiagnose verwenden zu können. Diese ist in Anlehnung an die sogenannten Kontrollkarten und verschiedene Prognosegütemaße aufgebaut.

Die Qualität des Verfahrens soll abschließend über unterschiedliche Kennzahlen durch Datensätze mit künstlich eingebauten Fehlern evaluiert werden und somit auch mit einer bereits bestehenden und an der LEC GmbH angewandten Methode verglichen werden. Diese beruht zwar auch auf den Modulen Modellfindung und Fehlerdiagnose, welche aber anders gestaltet sind. So erfolgt etwa die Modellbildung mit Hilfe der Hauptkomponentenzerlegung, auch Hauptkomponentenregression genannt.



## 2 Datenmaterial

Für die Modellbildung sowie für die Messdatendiagnose ist eine gewisse Datenbasis notwendig. Dieses Datenmaterial wird von der LEC GmbH zur Verfügung gestellt. Dabei handelt es sich um Aufzeichnungen zu unterschiedlichen Versuchen an einem Großgasmotor der Bauart Einzylinderforschungsmotor. Zu Beginn dieses Kapitels wird zuerst kurz erklärt, in welcher Form die Daten vorliegen. Die Auswahl an Messgrößen, die zur Analyse und Fehlerdiagnose herangezogen werden, wird in Abschnitt 2.2 gezeigt. Zu jeder Messreihe gibt es ein Protokoll, in dem beschrieben ist, welche Motorbauteile verwendet und welche Betriebspunkte im Versuch angefahren wurden. Dies wird in Abschnitt 2.3 behandelt. Eine Datenaufbereitung und -filterung wird ebenfalls durchgeführt, was vor jeglicher weiterer Verarbeitung der Daten, wie etwa zur Modellbildung, unumgänglich ist.

### 2.1 Datengrundlage

Die Datengrundlage bilden Messreihen, die am Forschungsmotor im Zeitraum vom 27. November 2014 bis zum 16. November 2015 gemessen bzw. aufgezeichnet wurden. Eine Messreihe ist die in zeitlicher Abfolge erfolgte Aufzeichnung der Messgrößen zu einem Messversuch am Prüfstand. In einem Versuch werden unterschiedliche Betriebspunkte angefahren, in dem etwa eine oder mehrere Größen variiert werden. Die Anzahl der Messpunkte, sprich die Länge der Messreihe, sowie der zeitliche Abstand zwischen ihnen sind von Messreihe zu Messreihe unterschiedlich. Die meisten der zugrundeliegenden Messreihen haben einen Umfang von 50 bis 200 Messpunkten. Der zeitliche Abstand dieser Punkte kann von einer Minute bis zu Messungen an unterschiedlichen Tagen variieren. Messreihen mit einer typischen Frequenz gibt es also nicht.

Zur Veranschaulichung folgen Beispiele von 8 Messzeitpunkten (Uhrzeiten im Format `hh:mm:ss`) in drei unterschiedlichen Messreihen. Messreihe 2 weist etwa Messungen im Minutentakt auf. Bei der dritten Reihe gibt es Messungen an zwei verschiedenen Tagen.

Messreihe 1	Messreihe 2	Messreihe 3
08:37:22	12:40:43	11:52:23 (1. Tag)
08:39:24	12:42:18	12:22:24
08:40:23	12:43:41	12:23:28
08:41:40	12:44:43	12:26:14
10:20:55	12:45:39	08:44:40 (2. Tag)
10:41:10	12:46:51	08:48:15
11:12:57	12:47:46	08:51:06
11:42:43	12:49:23	08:52:41

Ausgehend von 141 im genannten Zeitraum zur Verfügung stehenden Messreihen müssen zu Beginn Messreihen ausgeschlossen werden, bei denen nicht alle verwendeten Messgrößen aufgezeichnet wurden oder eine Aufzeichnung im geforderten Format sogar zur Gänze fehlt. Bei der anschließenden Analyse der Protokolle werden im Weiteren Messreihen entfernt, bei denen es im Protokoll Hinweise auf Unregelmäßigkeiten in den Messdaten gibt und bei denen spezielle Versuche angefahren wurden. Beispielsweise werden Messreihen mit einer Variation in der Drehzahl ausgeschlossen, da es zu wenig Messpunkte für statistisch tragfähige Aussagen geben würde.

Sämtliche gemessenen Größen sind in einer Datei im `ASCII`-Format aufgezeichnet. Zu Beginn der Analysen wird ein `.txt`-File erstellt, das die relevanten Messgrößen aller Messreihen enthält. Die Analyse der Daten, Modelle und Evaluierungsergebnisse wird mit der Statistiksoftware `R`<sup>®</sup>(Version 3.2.2) erledigt. Bei den Grafiken wird hier unter anderem das Paket `ggplot2` angewandt (siehe WICKHAM (2009)). Für die Modellauswahl, die Fehlerdiagnose und deren Evaluierung wird das Programmpaket `Matlab`<sup>®</sup>(Version R2012a) verwendet. Das Paket `R` ist frei zugänglich und vor allem für statistische Fragestellungen äußerst gut geeignet. `Matlab` hingegen ist eine kommerzielle Software und bietet große Vorteile bei numerischen Berechnungen. Weitere Informationen kann der Leser von der R Project Website ([www.r-project.org](http://www.r-project.org)), den MathWorks, Inc. ([de.mathworks.com](http://de.mathworks.com)) oder im Nachschlagewerk HIEBELER (2015) einholen.

## 2.2 Messgrößen

Insgesamt werden 23 Messgrößen für die Analyse herangezogen, welche in Tabelle 2.1 aufgelistet sind. Der Großteil davon beschreibt, wie in Abschnitt 1.1 erwähnt, Drücke, Massenströme, Temperaturen, Abgaskonzentrationen und motorspezifische Größen wie

etwa Drehzahl oder Drehmoment. Für 9 von diesen 23 Messgrößen sollen im Rahmen dieser Arbeit Modelle gebildet werden, die im Anschluss für eine Messdatendiagnose herangezogen werden. Bei diesen sogenannten Zielmessgrößen handelt es sich vor allem um Abgaskomponenten. Sie bilden den ersten Block in Tabelle 2.1 und sind **Rot** markiert.

Abkürzung	Maß	Beschreibung
mLuft	kg/h	Luftmassenstrom
mGas	kg/h	Kraftstoffmassenstrom
IMEP	bar	Indizierter Mitteldruck („Indicated Mean Effective Pressure“)
AbgO <sub>2</sub>	%	(Di)-Sauerstoff-Konzentration im Abgas
Lambda	-	Luftverhältnis gemessen mit Lambdasonde
AbgCO <sub>2</sub>	%	Kohlendioxid-Konzentration im Abgas
AbgCO	ppm	Kohlenmonoxid-Konzentration im Abgas
AbgNO <sub>x</sub>	ppm	diverse Stickoxide im Abgas
AbgHC	ppm	diverse unverbrannte Kohlenwasserstoffe im Abgas
mKW	kg/h	Massenstrom Kühlwasser
MZ	-	Methanzahl
xLuft	g/kg	Absolute Luftfeuchte
T <sub>2</sub>	°C	Lufttemperatur vor Motor
p <sub>2</sub>	bar	Luftdruck vor Motor
TKW <sub>vM</sub>	°C	Temperatur Kühlwasser vor Motor
aGasVK	%	Vorkammergasanteil
N	rpm	Drehzahl
M	Nm	Motordrehmoment
ZZP	$\frac{^{\circ}KW}{v_{OT}}$	Zündzeitpunkt
p <sub>3</sub>	bar	Abgasdruck
T <sub>3</sub>	°C	Abgastemperatur vor Gegendruckklappe
T <sub>4</sub>	°C	Abgastemperatur nach Gegendruckklappe
TKW <sub>nM</sub>	°C	Temperatur Kühlwasser nach Motor

Tabelle 2.1: Beschreibung der analysierten Messgrößen

Im Anschluss folgen jetzt zum besseren Verständnis detaillierte Beschreibungen einiger Messgrößen bzw. Maßeinheiten. Diese werden in den Werken VAN BASSHUYSEN et al. (Hrsg.) (2015) und BÖGE et al. (Hrsg.) (2015) noch ausführlicher beschrieben.

Bei den Massenströmen (Gas, Luft, Kühlwasser) können neben dem tatsächlichen Ausmaß unter anderem auch Druck, Temperatur und Feuchte gemessen werden. Die Me-

thanzahl bezeichnet ein Maß für die Klopfestigkeit von gasförmigen Kraftstoffen. Mit Klopfen bezeichnet man dabei Detonationen im Brennraum unterschiedlicher Ursache. Die Besonderheit an diesem Prüfstand ist die Verwendung einer Vorkammer. Da der Gasanteil im Hauptbrennraum für eine Zündung zu niedrig ist, passiert die Zündung nicht direkt im Brennraum, sondern wird in der Vorkammer ausgelöst.

Das Drehmoment, die Drehzahl und der Mitteldruck gehören zu den sogenannten Betriebskenngrößen eines Motors. Die Drehzahl beschreibt die Anzahl der Umdrehungen der Kurbelwelle je Zeiteinheit und wird deshalb in der Maßeinheit rpm angegeben. Dies ist die Abkürzung für revolutions per minute, auf Deutsch Umdrehungen pro Minute. Der innere oder indizierte Mitteldruck ist äquivalent zu der auf den Kolben wirkenden Arbeit. Er wird aus dem Zylinderdruckverlauf und dem Hubvolumen bestimmt. Der Zündzeitpunkt wird über °KW v. OT angegeben. °KW v. OT bedeutet Kurbelwinkel vor dem oberen Totpunkt, wobei als Totpunkte die Stellungen der Kurbelwelle eines Verbrennungsmotors bezeichnet werden, in denen der Kolben keine Bewegung mehr in axialer Richtung ausführt. Man unterscheidet zwischen dem oberen Totpunkt (Kolbenoberseite befindet sich nah am Zylinderkopf) und dem unterem Totpunkt (Gegenteil).

Lambda ( $\lambda$ ) bezeichnet das Luft-Kraftstoff-Verhältnis, welches die tatsächlich für eine Verbrennung zur Verfügung stehende Luftmasse ins Verhältnis zur mindestens notwendigen stöchiometrischen Luftmasse setzt, die für eine vollständige Verbrennung benötigt wird.  $\lambda > 1$  bedeutet „Luftüberschuss“, dh. mehr Luft nimmt an der Verbrennung teil als zur stöchiometrischen Reaktion notwendig ist. Die Lambdasonde vergleicht den Restsauerstoffgehalt im Abgas mit dem Sauerstoffgehalt in der Luft. Über die unterschiedlichen Abgaskomponenten und deren Entstehung kann in WIMMER (2012) nachgelesen werden. Dabei können nicht nur die Komponenten in ihren Anteilen sondern auch die Temperatur und der Druck des Abgases von Interesse sein.

### 2.3 Motorkonfiguration

Zu jeder Messreihe steht ein Protokoll zur Verfügung, in der die jeweilige Konfiguration des Motors festgehalten wurde. Eine Konfiguration bezeichnet dabei eine bestimmte Kombination von Bauteilen des Motors. Zu diesen austauschbaren Teilen gehören u.a. die Nockenwelle, der Kolben, die Vorkammer und die Zündkerze. Außerdem wird in den Protokollen noch festgehalten, welche Betriebspunkte am Prüfstand angefahren wurden. Dabei gibt es typische Folgen von Betriebspunkten, auch Variationen genannt, etwa der Betriebsmedien oder des Zündzeitpunktes. Im Folgenden werden die in der

Analyse betrachteten Bauteile aufgezählt und kurz beschrieben (siehe dazu auch BÖGE et al. (Hrsg.) (2015)).

Beim Kolben liegen Unterschiede in der Form und im Verdichtungsverhältnis  $\epsilon$  vor. Der Zylinderraum besteht aus Hubraum und Verdichtungsraum. Das Verdichtungsverhältnis gibt an, um wie viel mal der Zylinderraum größer ist als der Verdichtungsraum

$$\epsilon = \frac{V_h + V_c}{V_c} \quad \text{mit} \quad V_h + V_c = V_z,$$

wobei  $V_c$  den Verdichtungsraum,  $V_h$  den Hubraum und  $V_z$  den Zylinderraum bezeichnet. Die unterschiedlichen Formen werden über unterschiedliche Bezeichnungen deklariert. Bei der Nockenwelle gibt es Unterschiede im IVC (Abkürzung für „Intake Valve Closing“). Diese Kennzahl beschreibt den Schließzeitpunkt der Einlassventile. Beide Teile, Kolben und Nockenwelle, sind entscheidend für den Liefergrad. Dieser ist ein Maß für die im Zylinder nach Abschluss des Ladungswechsels verbleibende Frischladung. Bei der Vorkammer liegen Unterschiede im Volumen, in der Anzahl der Bohrungen und im Gasventil vor.

## 2.4 Datenaufbereitung

Aus physikalischen bzw. motortechnischen Gründen wurden die Messpunkte der Messreihen aus dem Datensatz ausgeschlossen, bei denen das Motordrehmoment negativ war. Dies ist bei sogenannten Schlepppunkten der Fall. In diesen Messpunkten ist die Zündung noch nicht aktiv. Der Motor wird daher von einem Elektromotor „geschleppt“. Weiters wird noch die C-Bilanz der Messpunkte betrachtet. Diese Bilanz ist zwar keine für die Analyse relevante Messgröße, dennoch können damit Messpunkte als unplausibel erklärt werden. Sie beschreibt die sogenannte Kohlenstoffbilanz als Verhältnis, in dem die abgeführte Menge an Kohlenstoffatomen in Relation zur zugeführten Menge gesetzt wird. Für eine nähere Beschreibung siehe dazu etwa FLOHR (2005). Diese Bilanz als Verhältnis sollte um 100% liegen. Ist dies nicht der Fall, so ist das ein Zeichen dafür, dass bei der Messung etwas nicht in Ordnung war, wie zum Beispiel das Vorliegen einer undichten Ansaugleitung. Auch andere Stoffbilanzen können zur Plausibilisierung der Messdaten herangezogen werden. Dies ist in diesem Fall jedoch nicht notwendig.

Weiters wird eine statistische uni- und bivariate Ausreißeranalyse durchgeführt. Hierbei werden sehr auffällige einzelne Ausreißer bzw. Gruppen von Ausreißern über die explorative Datenanalyse erkannt und (nach Rücksprache) entfernt. Dabei werden bei-

spielsweise Messpunkte aufgedeckt, bei denen für eine Messgröße keine Messung vorlag oder die Messgröße un plausible Werte (etwa negative) hatte. Das Ziel dieser Datenselektion ist es auch Punkte, die weit weg von der Masse der Daten sind, zu detektieren und gegebenenfalls auszuschließen, da diese keine Information über das typische Verhalten beinhalten und die Modelle dadurch nur verzerrt werden.

Am Ende der Datenaufbereitung stehen insgesamt 7945 Messpunkte von 82 Messreihen zur Verfügung. Dabei gibt es zu jedem Messpunkt und zu jeder Messgröße einen Wert. Es müssen in der Analyse somit keine sogenannten „missing values“ berücksichtigt werden. Wie in Abschnitt 4.6 ersichtlich sein wird, ist es sinnvoll die Messpunkte nach deren Motorkonfiguration, vor allem nach den Teilen Nockenwelle und Kolben, aufzuspalten. Gliedert man die zugrundeliegenden Messreihen nach diesen beiden Teilen (Nockenwelle nach IVC und Kolben nach Verdichtungsverhältnis  $\epsilon$  und Form), so sieht man, dass es dabei 3 unterschiedliche Nockenwellen und 4 unterschiedliche Kolben gibt. Da es für eine der 3 Nockenwellen nur Messreihen zu 2 der 4 möglichen Kolbentypen gibt, stehen also Messreihen zu 10 unterschiedlichen Kombinationen dieser 2 Bauteile zur Verfügung. Wie viele Messreihen und Messpunkte für jede dieser 10 Kombinationen zur Verfügung stehen, wird in Tabelle 2.2 ersichtlich.

Nockenwelle	Kolben	Konfiguration	Anzahl Messreihen	Anzahl Messpunkte
1	1	1	27	2113
	2	2	14	2307
	3	3	9	1064
	4	4	6	902
2	1	5	2	95
	2	6	4	479
	3	7	7	771
	4	8	2	111
3	1	9	4	35
	3	10	7	68

Tabelle 2.2: Messreihen aufgespalten nach den Bauteilen Kolben und Nockenwelle

Vor allem für Analysen zur Nockenwelle 3 stehen nur wenige Datenpunkte zur Verfügung. Die meisten Versuche wurden zur Nockenwelle 1 durchgeführt. Weiters ist darauf hinzuweisen, dass die Versuche der Konfigurationen 2 und 4 sogenannte Dauerläufe waren. Bei derartigen Versuchen ist das Ziel nicht die Variation von Betriebspunkten sondern die Überprüfung der Haltbarkeit gewisser Teile. Der Betriebspunkt wird da-

bei zu Beginn fix eingestellt. Durch Schwankungen der Randbedingungen streuen die Messgrößen zwar geringfügig aber nicht so stark, um für eine Modellanalyse geeignet zu sein. Bei der Explorativen Datenanalyse von Kapitel 3 werden sämtliche Messpunkte gemeinsam betrachtet. Bei der Modellbildung hingegen werden die Messpunkte nach ihrer Konfiguration unterschieden.



# 3 Explorative Datenanalyse

Vor jeder fortgeschrittenen statistischen Analyse ist es von Vorteil zuerst die Verteilung der vorliegenden Daten zu betrachten. Dazu werden vor allem Methoden der Explorativen Datenanalyse (EDA) angewandt. Dabei können sowohl gewisse Kenngrößen angegeben als auch verschiedene Grafiken verwendet werden. Die Analyse bezieht sich dabei sowohl auf eine uni- wie auch auf eine bivariate Datenanalyse. Die Datenbasis bilden dabei die gesamten 7945 Messpunkte ohne etwaige Unterscheidungen. Detailliertere Beschreibungen der deskriptiven Statistik können in den Werken ECKLE-KOHLER et al. (2009) oder CLEFF (2015) nachgelesen werden.

## 3.1 Univariate Datenanalyse

Für die univariate Analyse der Messdaten mittels statistischer Kenngrößen werden folgende Maße (Lagemaße, Streuungsmaße) verwendet, welche als Schätzer für die Momente über die Momentenmethode konstruiert sind bzw. als Schätzer für die theoretischen Quantile verwendet werden.

Für die Berechnung der Kenngrößen basierend auf Momente sei  $(x_1, \dots, x_n)$  eine Stichprobe aus metrischen Daten. Die Schätzer für die Momente um Null liefern u.a. diese Maße:

- Arithmetisches Mittel (Mittelwert)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

- Empirische Standardabweichung

$$sd = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

### 3 Explorative Datenanalyse

- Variationskoeffizient

$$cv = \frac{s}{\bar{x}}$$

Damit wird die Streuung der Daten als Vielfaches des Mittelwertes betrachtet, vorausgesetzt, dass  $x_i > 0$  für alle  $i = 1, \dots, n$  gilt.

Um die Kenngrößen basierend auf Quantile berechnen zu können, ist eine zumindest ordinal skalierte Stichprobe  $(x_1, \dots, x_n)$  notwendig. Weiters sei  $(x_{(1)} \leq \dots \leq x_{(n)})$  die Transformation dieser zu einer geordneten Stichprobe. Es können dann mittels der empirischen Quantile folgende Kennzahlen angegeben werden:

- Median

Man unterscheidet bei der Berechnung des Medians ( $x_{med}$ ) zwei Fälle:

$$n \text{ ungerade: } x_{med} = x_{\left(\frac{n+1}{2}\right)}$$

$$n \text{ gerade: } x_{med} = \frac{1}{2} \left( x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right)$$

- Minimum und Maximum

$$x_{min} = x_{(1)} \quad \text{und} \quad x_{max} = x_{(n)}$$

- $p$ -Quantil

Das  $p$ -te Quantil  $q_p$  der Stichprobe ist definiert durch

$$q_p = (1 - g)x_{(\lfloor (n-1)p \rfloor + 1)} + gx_{(\lfloor (n-1)p \rfloor + 2)},$$

wobei  $g = (n - 1)p - \lfloor (n - 1)p \rfloor$  und  $\lfloor p \rfloor$  die größte ganze Zahl kleiner gleich  $p$  angibt. Das Quantil  $q_p$  teilt den Bereich in  $(n - 1)$  Intervalle. Ein relativer Anteil von  $p$  der Intervalle ist links von  $q_p$  und ein Anteil von  $(1 - p)$  rechts davon. Weiters wird das  $q_{0.25}$ -Quantil als das 1. Quartil bezeichnet und das  $q_{0.75}$ -Quantil als das 3. Quartil.

Für die grafische Analyse werden Boxplots und Histogramme verwendet. Der Boxplot ist eine Summationsdarstellung der Verteilung der Daten durch 5 Kenngrößen, welche alle auf Quantile basieren. Die Höhe des Rechtecks bzw. der Box entspricht dem interquartilen Bereich ( $iqr = q_{0.75} - q_{0.25}$ ). Der horizontale Strich stellt den Median dar. Extremwerte oder sogenannte Ausreißer liegen außerhalb der Linien, deren Enden die um das 1,5-fache der Boxlänge verlängerten Quartile präsentieren. Beim Histogramm

sind die Häufigkeiten von den in disjunkte Klassen unterteilten Daten als Stäbe dargestellt.

In der folgenden Tabelle sind nun die Resultate der Berechnung der Kenngrößen zu den vorliegenden Messgrößen ersichtlich, wobei den ersten Block wiederum die 9 Zielgrößen bilden:

Messgröße	$x_{min}$	$q_{0.25}$	$x_{med}$	$\bar{x}$	$sd$	$cv$	$q_{0.75}$	$x_{max}$
mLuft	652.76	2643.35	2697.43	2735.22	357.65	0.13	2880.04	3724.64
mGas	22.10	76.41	77.46	77.79	8.52	0.11	82.04	94.85
IMEP	6.41	23.44	23.55	23.86	2.69	0.11	25.49	29.23
AbgO2	9.02	11.11	11.33	11.44	0.56	0.05	11.61	13.83
Lambda	1.69	2.04	2.09	2.11	0.11	0.05	2.14	2.70
AbgCO2	4.02	5.27	5.44	5.37	0.33	0.06	5.57	6.74
AbgCO	154.19	211.88	240.48	301.03	128.66	0.43	363.32	892.18
AbgNOx	11.61	51.99	130.28	106.91	52.28	0.49	137.35	855.82
AbgHC	134.17	412.34	502.30	756.62	506.59	0.67	1046.11	3372.26
mKW	10479.12	10502.69	11095.84	10887.78	286.92	0.03	11101.53	11125.53
MZ	70.09	85.68	88.22	87.68	2.77	0.03	90.01	95.75
xLuft	0.45	3.80	7.84	6.15	2.26	0.37	8.00	10.32
T2	34.80	50.00	50.20	57.63	8.33	0.14	65.00	75.60
p2	0.50	5.10	5.20	5.45	0.98	0.18	5.94	8.00
TKWvM	75.80	85.50	85.70	85.78	0.49	0.01	85.90	89.10
aGasVK	0.30	1.00	1.00	1.25	0.68	0.54	1.01	4.15
N	996.27	999.99	1000.00	999.96	0.26	0.00	1000.00	1000.31
M	842.99	4457.00	4537.38	4514.17	546.22	0.12	4805.85	5560.68
ZZP	6.00	20.00	20.00	19.90	1.79	0.09	20.00	30.00
p3	0.37	3.17	3.29	3.67	0.96	0.26	4.15	7.31
T3	447.09	531.59	552.79	545.34	23.21	0.04	559.64	607.08
T4	195.40	492.70	512.40	505.31	24.69	0.05	519.80	560.60
TKWnM	78.70	90.00	90.00	89.99	0.33	0.00	90.00	91.70

Tabelle 3.1: Deskriptive Kenngrößen aller Messgrößen,  $n = 7945$

Man sieht, dass die 9 Zielgrößen eine große Spannweite aufweisen. Hingegen ist diese bei manchen anderen Messgrößen, wie etwa N oder TKWnM, sehr gering. Weiters sind sämtliche Messgrößen im positiven Bereich. Ein besonderes Augenmerk wird hier auf die Betrachtung des Variationskoeffizienten, vor allem der Zielgrößen, gelegt. Dieser ist bei AbgHC, AbgNOx und AbgCO deutlich höher (0.43–0.67) als bei den anderen Zielgrößen ( $\leq 0.13$ ). Der Wertebereich und die Verteilung der Daten werden im Histogramm grafisch übersichtlicher dargestellt.

### 3 Explorative Datenanalyse

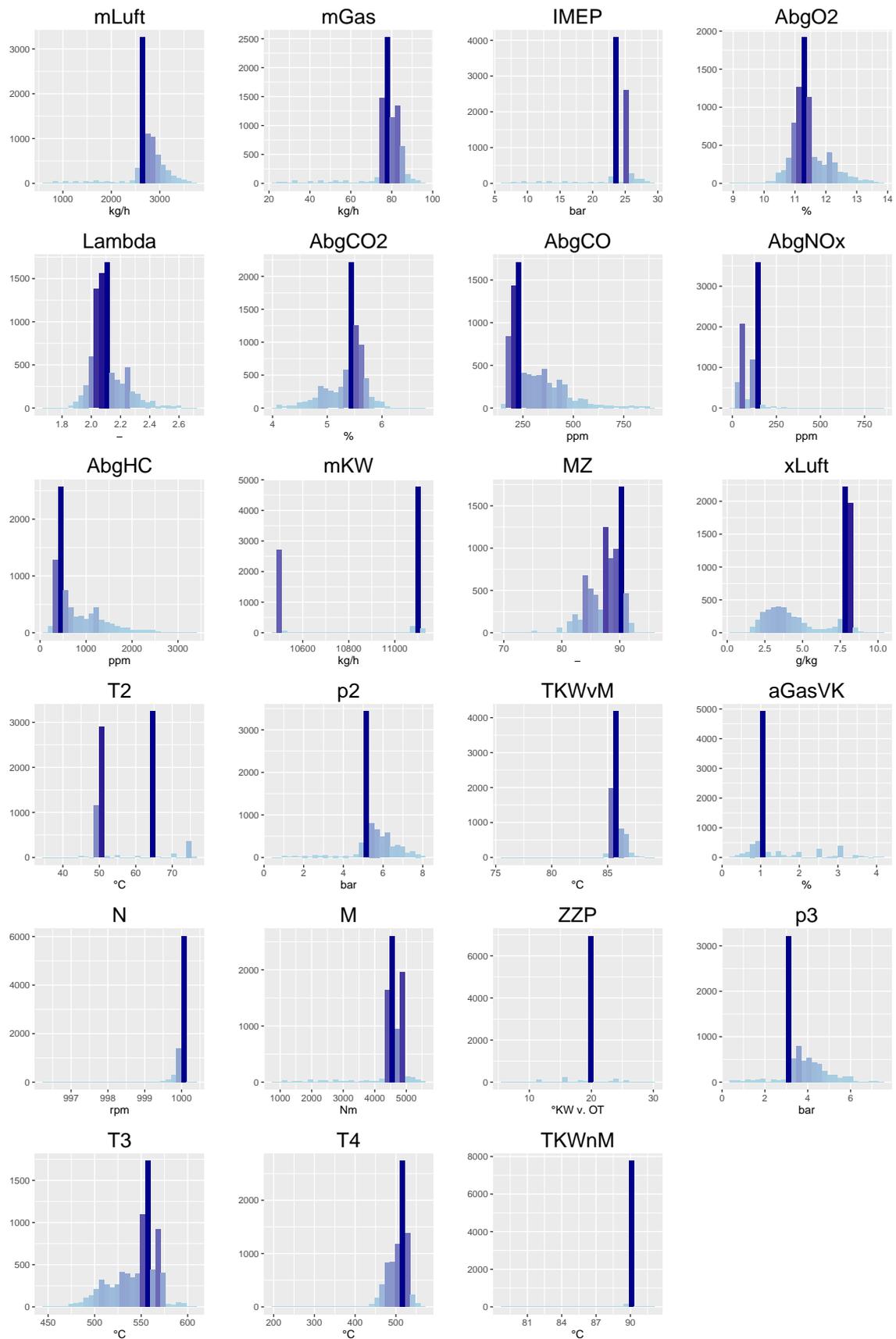
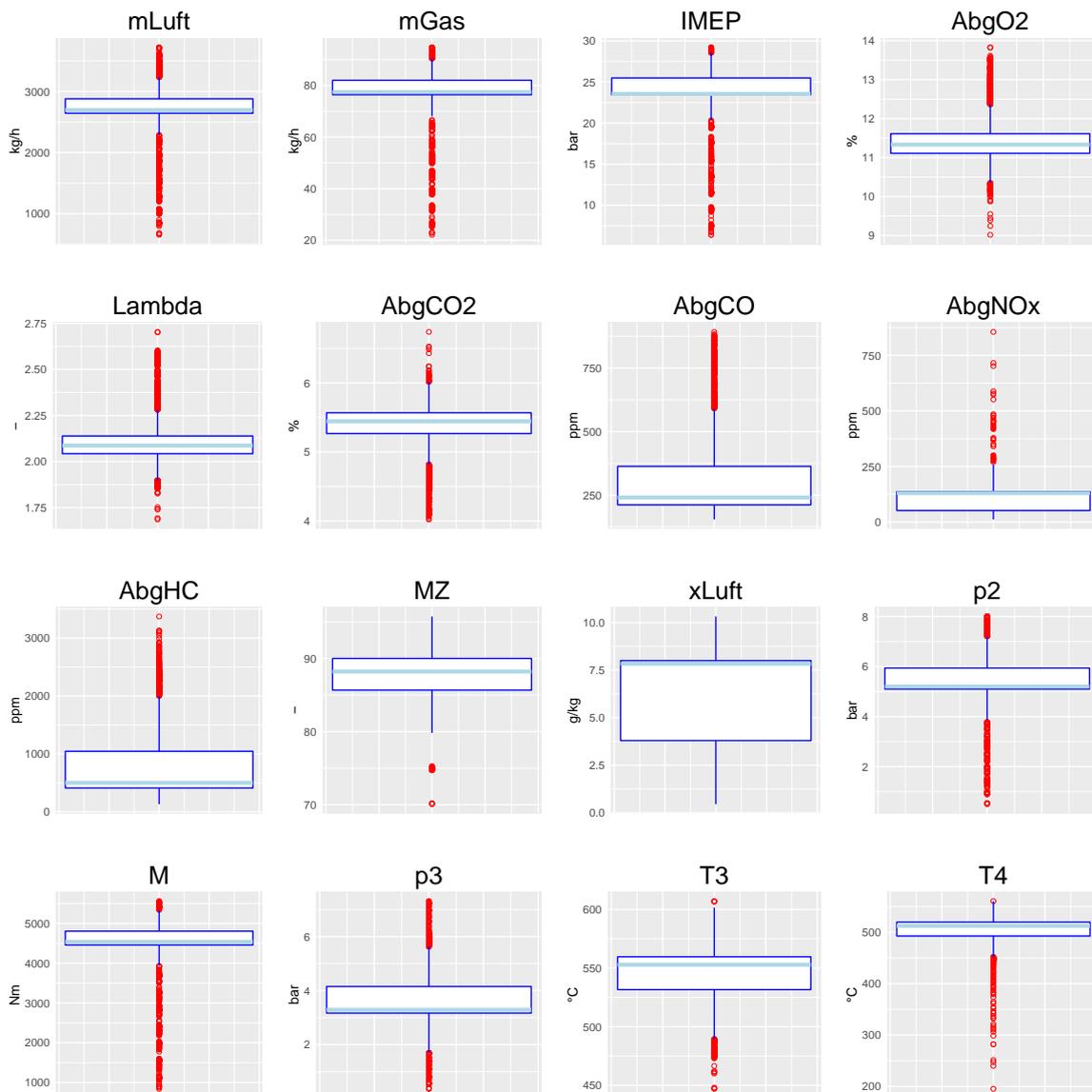


Abbildung 3.1: Histogramme aller Messgrößen,  $n = 7945$

Wie man sieht, gibt es einige Messgrößen deren Werte zum Großteil nur in 1 oder 2 Bereichen realisieren. Die Messgrößen Drehzahl  $N$ , Zündzeitpunkt  $ZZP$  und Temperatur des Kühlwassers  $TKWnM$  liegen meist auf einer Stufe. Die Lufttemperatur vor dem Motor  $p2$ , der Massenstrom des Kühlwassers  $mKW$  und der indizierte Mitteldruck  $IMEP$  sind auf 2 Stufen verteilt. Diese Messgrößen können als diskrete Variablen angesehen werden. Messgrößen mit 2 Stufen können aber sehr wohl als erklärende Größen verwendet werden, da es Unterschiede im Zusammenhang in den einzelnen Stufen geben kann. Bei Zielgrößen, wie etwa dem indizierten Mitteldruck, wird die Modellbildung dadurch erschwert. Für die Messgrößen, die mehr Variation aufweisen, werden nun die Boxplots dazu betrachtet.

Abbildung 3.2: Boxplots ausgewählter Messgrößen,  $n = 7945$

Es zeigt sich, dass die Zielgrößen einen sehr schmalen interquartilen Bereich im Vergleich zur Spannweite haben. Viele Messpunkte werden somit als Ausreißer markiert. Bis auf die Messgrößen MZ und `xLuft` trifft dies auch auf die restlichen Messgrößen zu.

## 3.2 Bivariate Datenanalyse

Da das Ziel die lineare Regression der Messgrößen ist, ist es sinnvoll den paarweisen Zusammenhang der Messgrößen zu betrachten. Diese bivariate Analyse kann wiederum wie bereits bei der univariaten über Kenngrößen und grafische Darstellungen geschehen. So können unterschiedliche Koeffizienten angegeben werden, die die Stärke des u.a. linearen Zusammenhangs ausdrücken. Als grafische Methode bieten sich sogenannte Scatterplots an, in denen die Punktepaaire in einem zweidimensionalen Koordinatensystem geplottet werden.

Sei  $(x_1, y_1), \dots, (x_n, y_n)$  eine bivariate Stichprobe. Für die Analyse des linearen Zusammenhangs wird der Korrelationskoeffizient von Pearson verwendet, der wie folgt berechnet wird:

$$r_P = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Des Weiteren kann der Zusammenhang mittels des Korrelationskoeffizienten von Spearman angegeben werden, der mittels Rangbildung nicht das Ausmaß eines linearen sondern eines monotonen Zusammenhangs misst. Dabei kann  $R(x_i)$  die Werte von 1 bis  $n$  annehmen und es gilt  $R(x_i) = j$ , falls  $x_i$  das  $j$ -größte Element der Stichprobe ist. Der Koeffizient wird somit folgendermaßen berechnet:

$$r_S = \frac{\sum_{i=1}^n \left( R(x_i) - \overline{R(x)} \right) \left( R(y_i) - \overline{R(y)} \right)}{\sqrt{\sum_{i=1}^n \left( R(x_i) - \overline{R(x)} \right)^2 \sum_{i=1}^n \left( R(y_i) - \overline{R(y)} \right)^2}}.$$

Beide Maße realisieren im Intervall  $[-1, 1]$ . Hohe Absolutwerte signalisieren dabei einen starken (linearen) Zusammenhang. Ein Wert um Null weist darauf hin, dass es kaum einen Zusammenhang der Größen gibt. In der folgenden Korrelationsmatrix sind beide Maße dargestellt, wobei die Größe und die Farbe der Punkte proportional zum Wert des jeweiligen Koeffizienten sind. Verwendet wird in R hierzu das Paket `corrplot` von WEI et al. (April 2016 - Version 0.77).

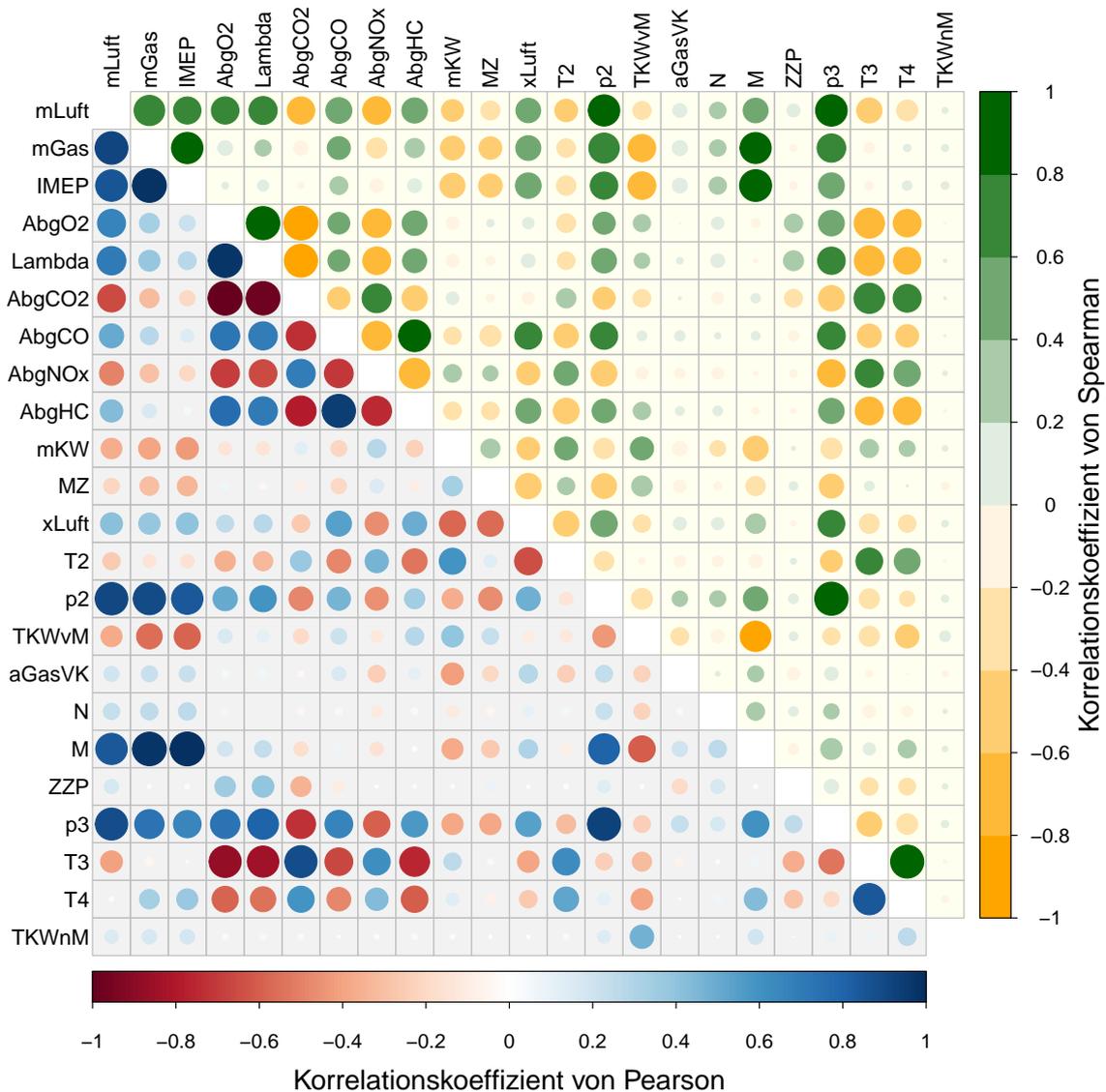


Abbildung 3.3: Korrelationsmatrix aller Messgrößen

Die Messgrößen N, ZZP und TKWnM stehen in keinem erkennbaren Zusammenhang zu anderen Messgrößen, was auch durch deren Verteilungseigenschaft (konzentriert auf einen Punkt) erklärbar ist. Hingegen kann man einen starken linearen Zusammenhang der Abgaswerte feststellen, die wiederum mit den Temperaturen des Abgases nach dem Motor sowohl vor als auch nach der Gegendruckklappe (T3 und T4) zusammenhängen. Eine weitere stark linear zusammenhängende Gruppe wird durch die Messgrößen mLuft, mGas, M, IMEP und p2 gebildet.

### 3 Explorative Datenanalyse

Für beide Gruppen werden nun noch einmal Scatterplotmatrizen erstellt, bei denen in der linken unteren Hälfte der Scatterplot zum jeweiligen Datenpaar ersichtlich ist und in der oberen rechten Hälfte der Korrelationskoeffizient von Pearson dargestellt ist. So sind der konkrete Wert des Koeffizienten und die tatsächliche Datenlage besser erkennbar

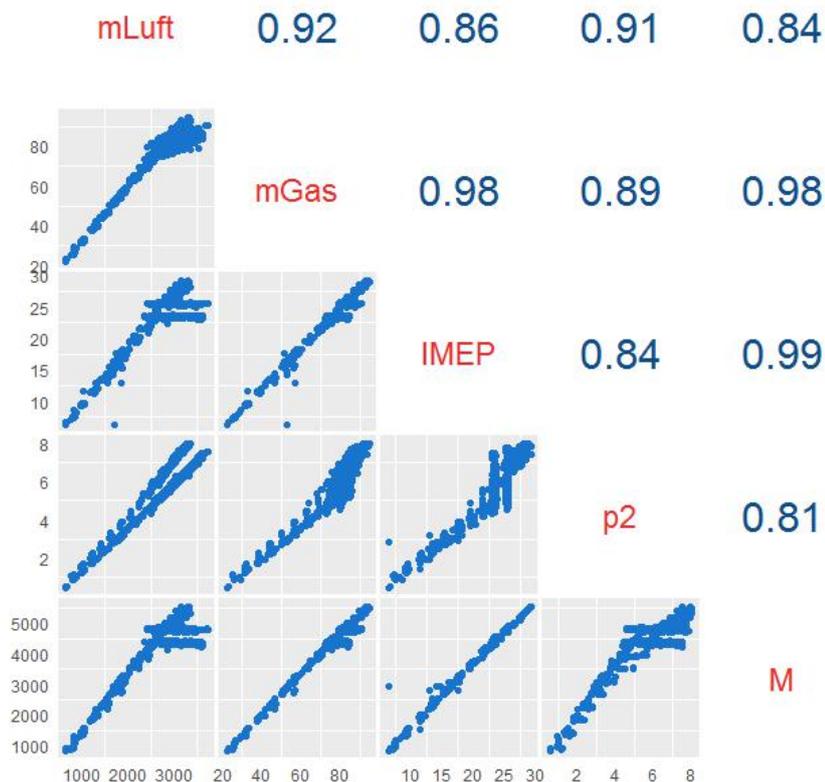


Abbildung 3.4: Scatterplotmatrix der Messgrößen mLuft, mGas, IMEP, p2 und M

Alle Paare weisen hier einen deutlichen linearen, und sogar positiven, Zusammenhang auf ( $r_P \geq 0.81$ ). Steigt eine Messgröße an, so steigen auch die anderen Messgrößen dieser Gruppe an. Der starke Zusammenhang der Zielgröße IMEP mit der Messgröße M ( $r_P = 0.99$ ) wird sich im dafür gewählten Modell widerspiegeln.

In der folgenden Abbildung wird ersichtlich sein, dass besonders die Abgaskomponenten AbgCO<sub>2</sub>, AbgO<sub>2</sub> und das Verbrennungsluftverhältnis Lambda einen signifikant hohen linearen Zusammenhang ( $|r_P| \geq 0.97$ ) haben. Dies ist auch eine logische Schlussfolgerung aus den physikalischen Eigenschaften der Verbrennung.

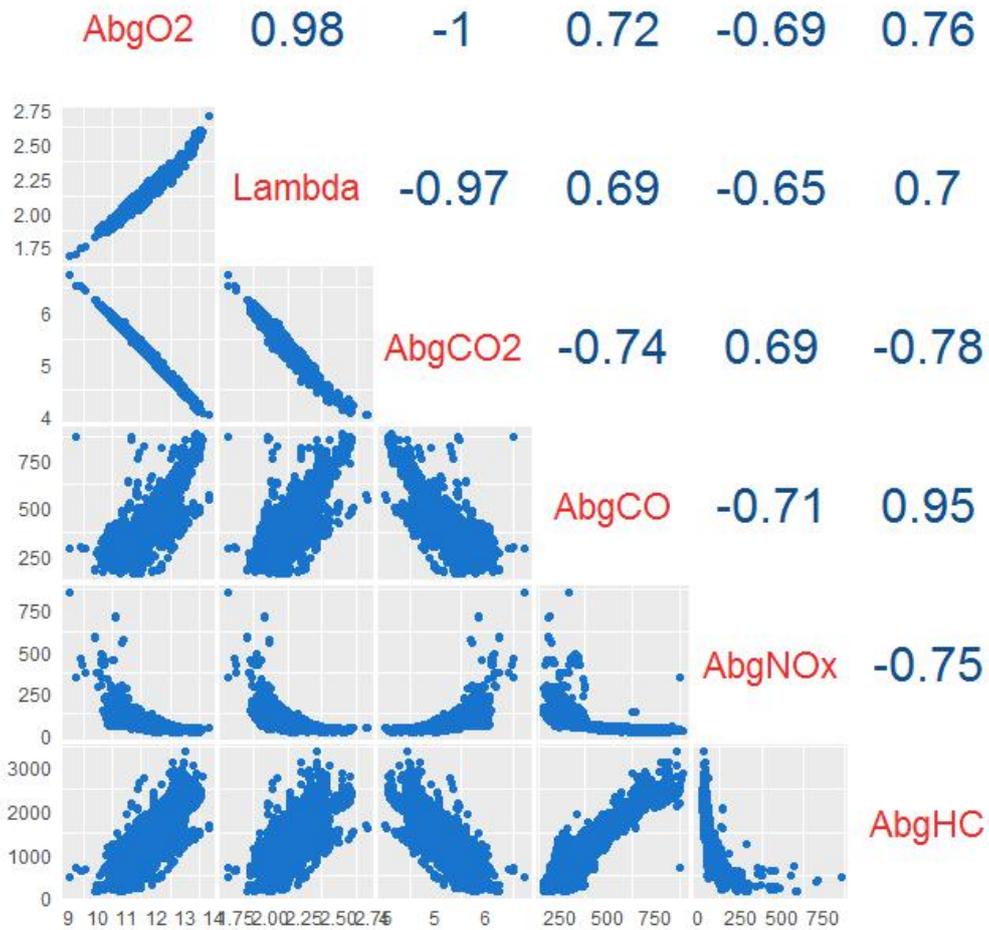


Abbildung 3.5: Scatterplotmatrix der Abgaskomponenten

Weiters interessant sind noch die Scatterplots mit der Messgröße  $AbgNO_x$ . Hier scheint der Zusammenhang zu den anderen Komponenten anhand der Scatterplots ersichtlich zwar gegeben zu sein, jedoch nicht in linearer Form ( $|r_P| \leq 0.75$ ). Nähere Details dazu folgen im nächsten Kapitel.



# 4 Modelle

Um für einen Messwert die Entscheidung zu treffen, ob dieser fehlerhaft ist oder nicht, kann man u.a. den tatsächlichen Messwert (Ist-Wert) mit einem Soll-Wert vergleichen. Im Rahmen dieser Arbeit erhält man diesen Sollwert über eine Prognose. Dabei werden auf Basis statistischer Modelle Vorhersagen gemacht bzw. Prognosewerte berechnet. Hierzu werden konkret multiple lineare Regressionsmodelle verwendet.

Deren theoretische Grundlagen werden beispielsweise im Buch von FAHRMEIR et al. (2009) behandelt und zu Beginn dieses Kapitels kurz erklärt. Ziel der Regressionsanalyse ist die Beschreibung einer Zielgröße mit Hilfe anderer Messgrößen. Da dafür jede Teilmenge vorhandener, erklärender Messgrößen verwendet werden kann, geht es darum mittels bekannter Auswahlverfahren Modelle zu finden, die bezüglich gewisser, festgelegter Kriterien möglichst gut sind. Sind die Modelle zunächst in ihrer Spezifikation gefunden, geht es dann darum die Modellparameter zu schätzen. Das Prinzip der Kleinsten-Quadrate-Schätzung findet dabei etwa Gebrauch.

Um auch qualitative Charakteristika in den Modellen zu berücksichtigen wird die Verwendung der sogenannten Indikatorvariablen kurz erläutert. Auf Grund der Annahmen, die in der linearen Regression getroffen werden, müssen die gewählten Modelle diesbezüglich evaluiert werden. Dies kann u.a. mittels Residuenplots geschehen. Abgeschlossen wird das Kapitel mit der praktischen Anwendung der erwähnten Theorie. So werden die Modelle zu den bereits erwähnten Zielgrößen vorgestellt und genauer analysiert.

## 4.1 Multiple lineare Regressionsmodelle

Ein Hauptziel der Regression ist es die Eigenschaften einer abhängigen Variable  $y$  (Response Variable) mittels Kovariablen  $x_1, \dots, x_k$  zu beschreiben. Die abhängige Variable wird unter anderem auch als Zielvariable und die Kovariablen als erklärende Variablen oder Prädiktoren bezeichnet.

Dabei ist der Zusammenhang der Zielgröße mit den Prädiktoren nicht exakt als Funktion gegeben, sondern durch zufällige Störungen überlagert. Die Zielgröße ist also eine Zufallsvariable, deren Verteilung von den erklärenden Variablen abhängig gemacht wird. Es wird nicht die Zielgröße direkt, sondern der (bedingte) Erwartungswert dieser in Abhängigkeit der Kovariablen modelliert. Dieser ist somit eine Funktion der Kovariablen

$$\mathbb{E}(y|x_1, \dots, x_k) = f(x_1, \dots, x_k).$$

Die Zielgröße lässt sich in Folge zerlegen in

$$y = \mathbb{E}(y|x_1, \dots, x_k) + \epsilon = f(x_1, \dots, x_k) + \epsilon.$$

Dabei wird  $\mathbb{E}(y|x_1, \dots, x_k) = f(x_1, \dots, x_k)$  als systematische Komponente und die zufällige Abweichung  $\epsilon$  als stochastische Komponente, auch Störgröße oder Fehlerterm bezeichnet. Ein Ziel der Linearen Regression besteht nun darin die systematische Komponente zu schätzen und von der stochastischen zu trennen.

Im Rahmen dieser Arbeit wird dabei die Klasse der linearen Regressionsmodelle verwendet, bei denen die Funktion  $f$  linear in den Parametern ist, so dass

$$\mathbb{E}(y|x_1, \dots, x_k) = f(x_1, \dots, x_k) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$$

gilt. Wenn nun also die Daten  $(x_{i1}, x_{i2}, \dots, x_{ik}, y_i)$  für jede Beobachtung  $i = 1, 2, \dots, n$  gegeben sind, ist ein multiples lineares Modell mit  $k$  erklärenden Variablen definiert durch die Modellgleichungen

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \epsilon_i, \quad i = 1, \dots, n,$$

wobei die Parameter  $\beta_0, \beta_1, \dots, \beta_k$  als Regressionskoeffizienten bzw.  $\beta_0$  als Intercept bezeichnet werden. Das Modell hat also  $p = k + 1$  unbekannte Parameter. Für die nicht beobachtbaren, statistischen Fehler wird angenommen, dass

$$\mathbb{E}(\epsilon_i) = 0 \quad \text{und} \quad \text{Var}(\epsilon_i) = \sigma^2$$

gilt und diese unabhängig verteilt sind. Die Eigenschaft gleich großer Varianzen dieser Fehler wird auch als Homoskedastizität bezeichnet. Zur Konstruktion von Konfidenzintervallen und Teststatistiken kann die Normalverteilungsannahme für  $\epsilon_i$  getätigt werden.

Die Formulierung kann dabei kompakter in Matrixnotation dargestellt werden. Definiert man etwa

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$

so können obige  $n$  Modellgleichungen zusammengefasst werden zu

$$y = X\beta + \epsilon,$$

wobei  $X$  eine  $n \times (k+1)$ -Matrix ist und als Modellmatrix oder Design-Matrix bezeichnet wird. Die Annahmen an die Zufallsfehler können formuliert werden als

$$\epsilon \sim N(0, I_n \sigma^2).$$

Das heißt:  $\mathbb{E}(\epsilon_j) = 0$  für alle  $j$ ,  $Cov(\epsilon_i, \epsilon_j) = \sigma^2$  für  $i = j$  und  $Cov(\epsilon_i, \epsilon_j) = 0$  für  $i \neq j$ . Als erste Konsequenz davon ergibt sich für die Verteilung von  $y$

$$\mathbb{E}(y) = \mu = X\beta \quad \text{und} \quad \text{Var}(y) = \sigma^2$$

bzw. mit Normalverteilungsannahme der Zufallsfehler sogar

$$y \sim N(X\beta, \sigma^2).$$

## 4.2 Kleinste-Quadrate-Schätzung

Man möchte nun die unbekannt Parameter  $\beta_0, \beta_1, \dots, \beta_k$  schätzen, um damit in Folge einen Schätzer für den Erwartungswert  $\mathbb{E}(y_i | x_1, \dots, x_k) = \mu_i$  zu erhalten:

$$\hat{\mu}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik}.$$

Für alle Beobachtungen  $i = 1, \dots, n$  kann dies folgendermaßen in Matrixnotation formuliert werden:

$$\hat{\mu} = X\hat{\beta}.$$

Eine Methode dafür ist die sogenannte Kleinste-Quadrate-Schätzung. Sie selbst und deren Ergebnisse werden anhand von SEN et al. (1990) kurz erläutert und sind im Werk noch detaillierter zu finden.

Das Prinzip der Kleinste-Quadrate-Schätzung ist die Quadratsumme der Residuen

$$SSE(\beta) = \sum_{i=1}^n (y_i - \mu_i)^2 = (y - X\beta)'(y - X\beta)$$

zu minimieren. Durch Differenzieren dieses Ausdruckes nach  $\beta$  und anschließendes Nullsetzen des Ergebnisses kann man zeigen, dass diese Residuenquadratsumme in

$$\hat{\beta} = (X'X)^{-1}X'y$$

ein Extremum hat. Es kann sogar bewiesen werden, dass  $SSE(\beta)$  in  $\hat{\beta}$  ein Minimum besitzt. Weiters kann gezeigt werden, dass  $\hat{\beta}$  ein erwartungstreuer Schätzer ist. Genauer gesagt gilt für dessen Erwartungswertvektor und Kovarianzmatrix:

$$\mathbb{E}(\hat{\beta}) = \beta \quad \text{und} \quad \text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}.$$

Das Einsetzen des Schätzers  $\hat{\beta}$  in obige Gleichung ergibt die Darstellung

$$\hat{\mu} = X\hat{\beta} = X(X'X)^{-1}X'y = Hy \quad \text{mit} \quad H = X(X'X)^{-1}X',$$

wobei  $H$  als Projektionsmatrix oder auch Hatmatrix bezeichnet wird. Diese Hatmatrix hat die Eigenschaft, dass sowohl sie selbst als auch  $(I - H)$  symmetrisch ( $H = H'$ ) und idempotent ( $H \cdot H = H$ ) ist. Dabei bezeichnet  $I$  die Einheitsmatrix, deren Einträge in der Diagonale 1 und ansonsten 0 sind. Für diesen Schätzer  $\hat{\mu}$  gelten folgende Verteilungseigenschaften:

$$\mathbb{E}(\hat{\mu}) = X \mathbb{E}(\hat{\beta}) = X\beta = \mu,$$

$$\text{Var}(\hat{\mu}) = \text{Var}(X(X'X)^{-1}X'y) = \sigma^2 X(X'X)^{-1}X'X(X'X)^{-1}X' = \sigma^2 X(X'X)^{-1}X'.$$

Die Residuen

$$r_i = y_i - \hat{\mu}_i$$

sind Schätzer für die nicht beobachtbaren Zufallsfehler  $\epsilon_i$  und definieren die Differenz zwischen den Beobachtungen und den geschätzten Erwartungswerten. Sie können mit Hilfe der Projektionsmatrix und der Einheitsmatrix folgendermaßen in Matrixnotation dargestellt werden

$$r = y - \hat{\mu} = y - Hy = (I - H)y.$$

Es können mit Hilfe der Projektionsmatrix und ihrer Eigenschaften in Folge interessante, geometrische Eigenschaften der Residuen hergeleitet werden.

So sind diese orthogonal zu den geschätzten Erwartungswerten  $\hat{\mu}$

$$\hat{\mu}'r = y'H(I - H)y = y'Hy - y'HHy = y'Hy - y'Hy = 0$$

und auch orthogonal zu jeder Spalte der Designmatrix

$$X'r = X'(I - H)y = X'y - X'Hy = X'y - X'X(X'X)^{-1}X'y = 0.$$

Außerdem folgt auch, dass die Summe der Residuen Null ist und dass für die Residuenquadratsumme und deren Erwartungswert in  $\hat{\beta}$  gilt:

$$SSE(\hat{\beta}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2 = r'r \quad \Rightarrow \quad \mathbb{E}(SSE(\hat{\beta})) = \sigma^2(n - p).$$

Somit ist  $SSE(\hat{\beta})/(n - p)$  ein erwartungstreuer Schätzer der unbekanntes Varianz  $\sigma^2$ .

Über die (geometrischen) Eigenschaften der Kleinsten-Quadrate-Schätzer kann in Folge auch eine Streuungszerlegungsformel für die empirische Varianz der Beobachtungen der Zielgröße hergeleitet werden, die wie folgt lautet:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{\mu}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{\mu}_i)^2.$$

Dabei bezeichnet

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2$$

die totale Quadratsumme, wobei  $SST/(n - 1)$  der empirischen Varianz der Beobachtungen entspricht,

$$SSR(\hat{\beta}) = \sum_{i=1}^n (\hat{\mu}_i - \bar{y})^2$$

die Regressions-Quadratsumme, welche den Unterschied zwischen dem Regressionsmodell und einem Intercept-only-Modell ( $\hat{\mu} = \bar{y}$ ) bewertet, und

$$SSE(\hat{\beta}) = \sum_{i=1}^n (y_i - \hat{\mu}_i)^2$$

die bereits bekannte Residuen-Quadratsumme, welche minimal im Kleinsten-Quadrate-Schätzer  $\hat{\beta}$  ist. Sämtliche gezeigten Folgerungen der Kleinste-Quadrate-Schätzung und ausführliche Beweise dazu kann der Leser in FRIEDL (2005) finden.

### 4.3 Modellselektion

Da es im Allgemeinen nicht das eindeutig beste Modell für eine Zielgröße gibt, liegt es in der Hand des Anwenders ein möglichst gutes Modell zu wählen. Die Güte des Modells kann dabei über unterschiedliche Kenngrößen bestimmt werden. Da für eine Menge an möglichen erklärenden Messgrößen jede Teilmenge davon als Prädiktoren eines Modells verwendet werden kann, ist es sinnvoll bereits bekannte Modellauswahlverfahren zu verwenden, um nicht sämtliche Modelle miteinander vergleichen zu müssen. Zu beachten ist dabei, dass diese Verfahren keine Garantie darauf geben das „beste“ Modell zu liefern, sofern es eines gibt. So können etwa unterschiedliche Auswahlverfahren zu unterschiedlichen Modellen führen.

Nach HACKL (2005) kann zur Beurteilung der Anpassungsgüte des Modells mit Hilfe der Streuungszerlegungsformel das sogenannte Bestimmtheitsmaß definiert werden

$$R^2 = \frac{SSR(\hat{\beta})}{SST(\hat{\beta})} = 1 - \frac{SSE(\hat{\beta})}{SST(\hat{\beta})}.$$

Dieses Maß hat jedoch den Nachteil, dass es bei Hinzunahme eines Prädiktors nicht geringer werden kann, egal ob dieser relevant ist oder nicht. Um dies zu kompensieren, kann eine Abwandlung des Bestimmtheitsmaßes verwendet werden, das sogenannte adjustierte Bestimmtheitsmaß

$$R_{adj}^2 = 1 - \frac{n-1}{n-p} \frac{SSE(\hat{\beta})}{SST(\hat{\beta})} = 1 - \frac{SSE(\hat{\beta})/(n-p)}{SST(\hat{\beta})/(n-1)}.$$

Wächst die Anzahl der Prädiktoren  $k = p - 1$  an, so wird der Faktor  $\frac{n-1}{n-p}$  größer und kompensiert damit, dass  $\frac{SSE(\hat{\beta})}{SST(\hat{\beta})}$  kleiner wird.

In der Literatur werden noch weitere Modellgütemaße diskutiert, wie etwa die von der logarithmierten Likelihood-Funktion im Schätzer  $\hat{\beta}$  abgeleiteten Informationskriterien: Das Akaike Informationskriterium und das Bayes'sche Informationskriterium. Der Vorteil dieser Kriterien ist, dass sie die Modellkomplexität über einen Strafterm berücksichtigen und so eher einfache Modelle vorgezogen werden, wenn man auf die Minimalität dieser Maße achtet. Für diese Zwecke wird, da die Anzahl an Parametern aus Gründen des Diagnoseverfahrens ohnedies möglichst klein gehalten werden sollte, in der Modellselektion das adjustierte Bestimmtheitsmaß als Kriterium herangezogen.

Kann man bereits anhand der Scatterplots starke lineare Zusammenhänge aufdecken, so können dadurch in Folge auf einfachem Weg lineare Modelle gefunden werden. Da

dies aber in den meisten Fällen nicht der Fall ist, gibt es unterschiedliche Methoden ein „gutes“ Modell zu finden. Im Rahmen dieser Arbeit wird die sogenannte Schrittweise Regression angewandt, bei der die Prädiktoren Schritt für Schritt ausgewählt werden. Das Schema dieses Vorgehens ist unter anderem in KLEINBAUM et al. (1998) zu finden und lautet wie folgt:

- Wähle ein Modellgütemaß zur Modellevaluierung und bestimme die Prädiktoren, die möglicherweise ins Modell aufgenommen werden sollen, sprich ein maximales Modell.
- Starte mit einem Initialmodell, das entweder nur den Intercept oder bereits auch schon gewisse Prädiktoren enthält.
- In jedem Schritt wird im Hinblick auf das Modellgütemaße entschieden, ob ein Prädiktor in das bestehende Modell aufgenommen oder ein Prädiktor daraus entfernt wird. Eine erklärende Größe, die in einem früheren Schritt aufgenommen wurde, kann auf Grund ihrer Zusammenhänge mit den anderen überflüssig geworden sein. Ergebnis eines jeden Schrittes soll ein Modell mit einem besserem Modellgütemaß als im Schritt zuvor sein. Bei mehreren Möglichkeiten wird der Schritt mit der größten Verbesserung gewählt.
- Man stoppe dann, wenn die Hinzu- bzw. Wegnahme von Prädiktoren keine signifikante Änderung, sprich eine Verbesserung im Modellgütemaß, mehr ergibt.

In `Matlab` kann dies unter anderem via `LinearModel.stepwise` durchgeführt werden. Dabei können neben dem Initialmodell und der gesamten Prädiktorenmenge auch noch viele weitere Einstellungen getroffen werden, etwa auch wie stark die Verbesserung im Modellgütemaß sein muss, um einen Prädiktor hinzuzufügen oder rauszunehmen. In `R` steht dafür der Befehl `stepAIC` zur Verfügung. Hier stehen die verschiedenen, zuvor erwähnten Informationskriterien als Kriterium für die Modellwahl zur Verfügung.

Weiters soll darauf hingewiesen werden, dass sich die Linearität des Modells nur auf dessen Parameter  $\beta_j$  bezieht. Etwaige nicht-lineare Transformation der Prädiktoren  $x_j$  oder die Einbindungen polynomieller Terme sind durchaus möglich.

## 4.4 Indikatorvariablen

Die Daten stammen wie in Abschnitt 2.3 beschrieben aus Messreihen zu unterschiedlichen Konfigurationen am Prüfstand. Gliedert man diese etwa nach den Bauteilen

Nockenwelle und Kolben auf, so sieht man, dass es 10 verschiedene Kombinationen dieser Teile gibt, wie auch in Tabelle 2.2 ersichtlich. Um die daraus entstehenden, sogenannten Strukturbrüche zu berücksichtigen, können Indikatorvariablen ins Modell aufgenommen werden. Es soll hierbei kein Unterschied in der Spezifikation des Modells durch die Prädiktoren, sondern in den Parametern aufgezeigt werden.

Dass es so einen Unterschied zwischen den Konfigurationen gibt, kann man zuallererst auch grafisch aufdecken, in dem man sich den Residuenplot, bei dem die standardisierten Residuen gegen die Vorhersagewerte geplottet werden (siehe Abschnitt 4.5), für das Gesamtmodell über alle Messpunkte ansieht und gewisse Strukturen in den Datenpunkten entdeckt, die bei normalen Verhalten ein willkürliches Muster aufweisen sollten. Besonders deutlich können die unterschiedlichen Strukturen aufgedeckt werden, wenn man die Punktepaare des Scatterplots nach den unterschiedlichen Konfigurationen einfärbt.

Durch die Verwendung sogenannter Indikator-Variablen bzw. Modelle mit Indikator-Variablen können auch solche Analysen von Strukturbrüchen passieren. Dies wird etwa im Werk KLEINBAUM et al. (1998) folgendermaßen beschrieben: Eine erklärende Größe, die eine endliche Anzahl an Werten hat, sodass unterschiedliche Kategorien einer Größe mit nominaler Messskala beschrieben werden können, nennt man Indikator-, Schein- oder Dummy-Variable. Die Regressionsanalyse funktioniert auch mit solchen kategorischen Variablen. Dadurch kann der Einfluss der in Kategorien realisierenden Variable qualitativ beschrieben werden.

Grundsätzlich muss dabei die einfache Regel beachtet werden, dass für eine erklärende Größe mit  $k$  Kategorien genau  $k - 1$  Indikatorvariablen eingeführt werden müssen, so fern das Modell einen konstanten Term ( $\beta_0$ ) enthält. Geht man etwa beispielsweise von einem Modell mit 2 Prädiktoren der Form

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \epsilon$$

aus und möchte man eine Charakteristik mit drei Stufen miteinbeziehen, dann benötigt man dafür zwei Indikatorvariablen  $z_1$  und  $z_2$  und erweitert das Modell zu:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot z_1 + \beta_4 \cdot z_2 + \beta_5 \cdot z_1 \cdot x_1 + \beta_6 \cdot z_2 \cdot x_1 + \beta_7 \cdot z_1 \cdot x_2 + \beta_8 \cdot z_2 \cdot x_2 + \epsilon$$

$$\text{mit } z_i = \begin{cases} 1 & \text{wenn Datenpunkt von Stufe } i, \\ 0 & \text{sonst.} \end{cases} \quad \text{für } i = 1, 2.$$

Über diese Definition wird Stufe bzw. Kategorie 3 der nominalen Variable als Referenz gewählt. Auch andere Referenzstufen und Kodierungsweisen, etwa mit  $-1$  und  $1$  statt  $0$  und  $1$ , sind möglich.

Durch die Modellgleichungsformel wird ersichtlich, dass mit Hilfe der Indikatorvariablen für jede Stufe der nominalen Variable eine Verschiebung im Intercept und, auf Grund der Wechselwirkung mit den beiden erklärenden Größen, eine Änderung in den Zusammenhängen mit den Prädiktoren  $x_1$  und  $x_2$  möglich ist. Es ist weiters möglich mehrere qualitative Charakteristika in ein Modell miteinzubeziehen, und auch deren Wechselwirkung kann analysiert werden. Näheres dazu findet der Leser etwa in KLEINBAUM et al. (1998).

## 4.5 Modellvalidierung

Die Modellvalidierung bzw. Überprüfung der Modellannahmen kann unter anderem durch so genannte Diagnoseplots geschehen (siehe dazu STADLOBER (2011)). Darin wird vor allem die Anwendung von Residuenplots angeführt. Die Residuen sind wie bereits erwähnt gegeben durch

$$r_i = y_i - \hat{\mu}_i.$$

Sie sind Schätzer der nicht beobachtbaren Zufallsfehler  $\epsilon_i \stackrel{iid}{\sim} N(0, \sigma^2)$ . Im Zuge der Modellvalidierung kann man auch die standardisierten Residuen betrachten

$$r_i^* = \frac{r_i}{\sqrt{MSE(\hat{\beta})}} \quad \text{mit} \quad MSE(\hat{\beta}) = \frac{SSE(\hat{\beta})}{n - p},$$

die Varianz  $\approx 1$  besitzen.

Die Normalverteilungsannahme wird durch einen Q-Q-Plot bzgl. der beobachteten (standardisierten) Residuen grafisch getestet. Bei diesem Quantil-Quantil-Plot werden die empirischen Quantile der Residuen gegen die theoretischen Quantile der Normalverteilung geplottet. Die Referenz ist ein lineares Punktemuster. Weicht das Muster von einer Geraden ab, so deutet dies auf Verletzungen der Normalverteilungsannahme hin.

Durch einen Scatterplot, bei dem die standardisierten Residuen  $r_i^*$  gegen die Vorhersagewerte  $\hat{\mu}_i$  geplottet werden, kann die Annahme der konstanten Varianz überprüft werden. Außerdem kann damit die Unabhängigkeit der Residuen von den Vorhersa-

gewerten grafisch getestet werden. Eine Punktwolke mit zufälliger Struktur ohne erkennbares Muster bildet hier die Referenz. Trichterförmige Muster weisen etwa auf eine Zu- bzw. Abnahme der Varianz in Abhängigkeit der Vorhersagewerte hin.

## 4.6 Resultate der praktischen Anwendung

Auf Basis der zuvor beschriebenen theoretischen Grundlagen werden nun für die vorgestellten 9 Zielgrößen unter anderem mit Hilfe der Schrittweisen Regression Modelle gebildet, für die Modelle entscheidende Bauteile des Motors mit Hilfe der Indikatorvariablen aufgezeigt, und zum Abschluss die Annahmen zu den gefundenen Modellen überprüft. Weiters wird auf den Fall der Zielgröße **AbgNO<sub>x</sub>** eingegangen. Hier ist es vorteilhaft eine Transformation der Prädiktoren durchzuführen. Sämtliche ausgewählten Modelle und Details dazu sind in Tabelle 4.2 aufgelistet.

### 4.6.1 Modellfindung

Bei der Modellfindung wird so vorgegangen, dass man die Modelle mit Teilen der Gesamtdaten fittet, versucht ein gemeinsames Modelle für alle Teilmengen zu finden und anschließend die Güte des Modells überprüft. Die Teilung der Daten wird hier durch vorgegebene Konfigurationen am Prüfstand vorgenommen. Andere zufällige Teilungen zeigen dabei keine zufriedenstellenden Ergebnisse, da in manchen Modellen der Unterschied in den Konfigurationen über Messgrößen zu erklären versucht wird. Außerdem hat man den Vorteil, dass damit gleichzeitig überprüft werden kann, ob dieselben Modelle in allen Datenpunkten, wenn auch mit unterschiedlichen Schätzungen, gute Anpassungen liefern. Wie bereits in Abschnitt 4.4 erwähnt sollen die Modelle in allen Datenpunkten gelten bzw. sich bei Unterschieden in den Datenpunkten durch qualitative Charakteristika, wie etwa Motorbauteile, nicht in ihrer Spezifikation sondern nur durch die Schätzungen der unbekannt Parameter unterscheiden.

Die genaue Vorgehensweise ist eine Kombination der Anwendung der vorgestellten Schrittweisen Regression und einer händischen Verbesserung der damit erhaltenen Ergebnisse durch Vergleich und Anpassung der Modelle. Sie wird in Abbildung 4.1 schematisch dargestellt. Wichtig für die Modellbildung ist in diesem Fall, dass die Modelle, wenn möglich, ohne andere Zielgrößen als erklärende Größen gebildet werden sollen. Außerdem soll die Anzahl an erklärenden Prädiktoren so gering wie möglich gehalten werden, um zusätzliche Fehlerquellen zu vermeiden. Besonders in Hinsicht darauf bedarf es einer Überprüfung der Ergebnisse der Schrittweisen Regression, da diese unter

Umständen eine zu große Anzahl an Prädiktoren vorschlägt.

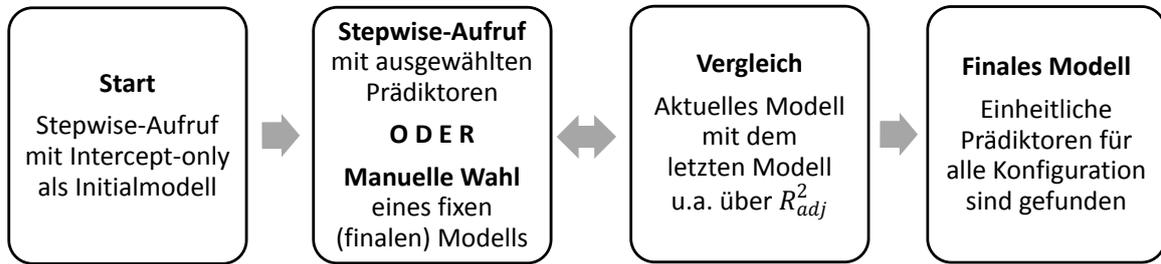


Abbildung 4.1: Schemahafte Darstellung der Modellfindung

Das prinzipielle Vorgehen wird in Folge anhand der Modellfindung für die Messgröße `AbgC02` und für die beiden Konfigurationen mit den meisten Datenpunkten und ohne Dauerlaufversuche (`Konfiguration 1` und `Konfiguration 3`) erklärt. Dabei wird das finale Modell in 3 Schritten gefunden: Schrittweise Regression mit dem Intercept-only-Modell ( $\hat{\mu} = \hat{\beta}_0 = \bar{y}$ ) als Initialmodell, Schrittweise Regression mit manuell ausgewählten Prädiktoren im Initialmodell und finale Auswahl der Prädiktoren. Die Ergebnisse nach dem 1. Schritt, der Durchführung des `LinearModel.stepwise`-Befehls in `Matlab`, für die Messgröße `AbgC02` zu den Messpunkten der beiden Konfigurationen sehen folgendermaßen aus:

#### Konfiguration 1

1. Adding p3, AdjRsquared = 0.91172
2. Adding T3, AdjRsquared = 0.96359
3. Adding ZP, AdjRsquared = 0.97257
4. Adding ZP^2, AdjRsquared = 0.97728
5. Adding p3:T3, AdjRsquared = 0.9802

Linear regression model:

$$y \sim 1 + ZP + p3*T3 + ZP^2$$

Number of observations: 2113

Error degrees of freedom: 2107

Root Mean Squared Error: 0.0504

R-squared: 0.98

Adjusted R-Squared 0.98

#### Konfiguration 3

1. Adding p3, AdjRsquared = 0.90434
2. Adding T3, AdjRsquared = 0.97182
3. Adding T4, AdjRsquared = 0.98077
4. Adding M, AdjRsquared = 0.9837
5. Adding p2, AdjRsquared = 0.98941
6. Adding p2^2, AdjRsquared = 0.99598
7. Removing T4, AdjRsquared = 0.99594

Linear regression model:

$$y \sim 1 + p2 + M + p3 + T3 + p2^2$$

Number of observations: 1064

Error degrees of freedom: 1058

Root Mean Squared Error: 0.0223

R-squared: 0.996

Adjusted R-Squared 0.996

Wie man sofort erkennen kann, unterscheiden sich die beiden durch die Schrittweise Regression automatisch erstellten Modelle in ihrer Spezifikation. Außerdem wird anhand der Konfiguration 3 auch der Schritt der Stepwise Regression erkenntlich, bei dem ein Prädiktor aus dem Modell genommen wird. Die Regressionsanalyse aller 10 Konfigurationen (siehe Tabelle 4.1) lieferte, wie zu erwarten war, unterschiedlich spezifizierte Modelle. Das Ziel ist es aber für alle Konfigurationen brauchbare Modelle mit gleicher Spezifikation zu finden.

Da es im Allgemeinen nicht das eindeutig beste Modell gibt, kann es durchaus sein, dass sich die erklärenden Prädiktoren der einzelnen Konfigurationen unterscheiden, es aber trotzdem ein Modell gibt, das in allen Konfigurationen ein ähnlich hohes Bestimmtheitsmaß erreicht. Da sich die Prädiktorenmengen der Modelle je Konfiguration mittels der Schrittweisen Regression unterscheiden (können), muss nun ein gemeinsames Set an Prädiktoren gefunden werden. Hilfreich ist dabei, dass im Befehlsaufruf auch ein Initialmodell mit erklärenden Größen angegeben werden kann.

Die Ergebnisse des 1. Schrittes für alle zehn Konfigurationen in Tabelle 4.1 enthielten in den meisten Fällen die Messgrößen `p3`, `T3` und `p2` als Prädiktoren. Der nächste Versuch bzw. der 2. Schritt ist also mit einem Initialmodell zu starten, das die Prädiktoren `p3`, `T3` und `p2` enthält:

Konfiguration 1

1. Adding M, AdjRsquared = 0.98918
2. Adding `p3^2`, AdjRsquared = 0.99271

Linear regression model:

$$y \sim 1 + p2 + M + p3 + T3 + p3^2$$

Number of observations: 2113

Error degrees of freedom: 2107

Root Mean Squared Error: 0.0306

R-squared: 0.993

Adjusted R-Squared 0.993

Konfiguration 3

1. Adding M, AdjRsquared = 0.98663
2. Adding `p2^2`, AdjRsquared = 0.99594

Linear regression model:

$$y \sim 1 + p2 + M + p3 + T3 + p2^2$$

Number of observations: 1064

Error degrees of freedom: 1058

Root Mean Squared Error: 0.0223

R-squared: 0.996

Adjusted R-Squared 0.996

Obwohl sich die Prädiktorenmengen der Modelle je Konfiguration mittels der Stepwise-Prozedur nun noch immer unterscheiden, wird trotzdem ein Versuch gestartet ein Zielmodell anzugeben. Die Modelle mit den resultierenden Prädiktoren sollen dabei

für jede Konfiguration ein annähernd gleich großes Bestimmtheitsmaß erreichen wie die Modelle, die durch die Schrittweise Regression vorgeschlagen wurden. Betrachtet man wiederum sämtliche Ergebnisse dieses Zwischenschrittes, so werden zum Initialmodell mit den Messgrößen p3, T3 und p2 noch die Messgrößen T4 und M hinzugenommen. Für die Zielgröße Abgas C02 sehen die Ergebnisse dieses finalen Modells für beide Konfigurationen folgendermaßen aus:

## Konfiguration 1

Linear regression model:

$$y \sim 1 + p2 + M + p3 + T3 + T4$$

Estimated Coefficients:

	Estimate	tStat
(Intercept)	2.3607	45.902
p2	-0.8268	-68.516
M	0.00071659	67.799
p3	0.2493	29.978
T3	0.0070634	45.85
T4	-0.0012138	-8.6191

Number of observations: 2113

Error degrees of freedom: 2107

Root Mean Squared Error: 0.0366

R-squared: 0.99

Adjusted R-Squared 0.99

## Konfiguration 3

Linear regression model:

$$y \sim 1 + p2 + M + p3 + T3 + T4$$

Estimated Coefficients:

	Estimate	tStat
(Intercept)	2.0739	18.028
p2	-0.67054	-23.921
M	0.0006291	29.133
p3	0.19448	12.924
T3	0.0093565	28.73
T4	-0.0035499	-16.712

Number of observations: 1064

Error degrees of freedom: 1058

Root Mean Squared Error: 0.0359

R-squared: 0.989

Adjusted R-Squared 0.989

Es zeigt sich, dass auch das neue Modell, mit zum Teil anderen Prädiktoren als durch die Schrittweise Regression vorgeschlagen, ein ähnlich hohes Bestimmtheitsmaß für beide Konfigurationen liefert und es somit auch ein Modell gibt, dass in seiner Spezifikation in beiden Messpunktgruppen gültig ist, auch wenn sich die Schätzungen leicht unterscheiden.

Wie bereits erwähnt wird das obige Vorgehen nicht nur über die Betrachtung der Ergebnisse von 2 Konfigurationen sondern bezüglich der Analyse zu allen 10 Konfigurationen durchgeführt. Die Ergebnisse der jeweiligen Schritte sind dabei in der folgenden Tabelle aufgeführt. Dabei sind sowohl die Ergebnisse der Stepwise Regression (Modellformel ohne Interceptterm) ersichtlich als auch die Modellgüte der damit gefundenen Modelle.

Weiters ist das Modellgütemaß des finalen Modells, in **Rot** markiert, abzulesen. Insbesondere fallen die Konfigurationen 2 und 4 auf, bei deren Versuche es sich wie bereits in Abschnitt 2.4 erwähnt um Dauerläufe handelt. Aufgrund der geringen Variation der Zielgrößen und auch der erklärenden Größen wird hier eine auffällig große Anzahl an Prädiktoren verwendet. Das finale Modell mit deutlich weniger Zielgrößen weist deshalb ein geringeres Modellgütemaß aus. Ansonsten scheint das finale Modell in allen Konfigurationen einen zufriedenstellenden Anpassungsgrad zu bieten.

K	1. Stepwise Ergebnis	$R_{adj}^2$	2. Stepwise Ergebnis	$R_{adj}^2$	$R_{adj}^2$
1	$ZZP + p3 * T3 + ZZP^2$	0.98	$p2 + M + p3 + T3 + p3^2$	0.99	0.99
2	$T2 + p2 + T4 + xLuft * M$ $+ M * T3 + p2^2 + M^2$	0.96	$T4 + xLuft * p3 + p2 * T3$ $+ M * p3 + p3^2$	0.96	0.84
3	$p2 + M + p3 + T3 + p2^2$	0.99	$p2 + M + p3 + T3 + p2^2$	0.99	0.99
4	$p2 + N + M + p3 + T3$ $+ T4 + xLuft * MZ + MZ^2$	0.82	$p2 + N + M + p3 + T3$ $+ T4 + xLuft * MZ + MZ^2$	0.82	0.69
5	$p2 + M + p3 + T2 * T4 + p3^2$ $+ T2 * aGasVK + xLuft * T3$	0.98	$p2 + T4 + xLuft * T3 + p2^2$ $+ T2 * M + T2 * aGasVK$	0.98	0.81
6	$T2 + M + p3 + MZ + p3^2$	0.99	$T2 + M + p3 + T3 + p3^2$	0.99	0.98
7	$M + p3 + p3^2$	0.99	$T2 + p2 + M + p3 + T3 + T4$	0.99	0.99
8	$p3 + T3 + T4 + TKWvM * MZ$ $+ TKWvM^2 + p3^2 + T3^2$	0.99	$p2 + M + p3 + T3$ $+ TKWnM + T4 + T4^2$	0.99	0.97
9	$T2 + p2 + ZZP + p3 + T3$ $+ T4 + p2^2$	0.99	$T2 + p2 + ZZP + p3 + T3$ $+ T4 + p2^2$	0.99	0.96
10	$M + ZZP + aGasVK + p2 * T3$	0.99	$ZZP + T3 + p2 * p3 + M * p3$	0.99	0.97

Tabelle 4.1: Beispielhafte Modellfindung anhand der Zielgröße AbgC02

Betrachtet man die Tabelle genauer so wird ersichtlich, dass mit diesem Verfahren auch Wechselwirkungen der Prädiktoren und polynomielle Terme in den Modellen in Frage kommen. Betrachtet man die Unterschiede zwischen den beiden Schritten so fällt auf, dass, falls die manuell ausgewählten Prädiktoren im 1. Schritt schon vorhanden waren, sich die Modelle im 2. Schritt nicht unterscheiden, wie etwa bei Konfiguration 3. Bei manchen Konfigurationen werden aber im Zuge der automatischen Modellwahl auch die Initialprädiktoren aus dem Modell gegeben. Die Größe p2 ist zum Beispiel in Konfiguration 6 im Modell vom 2. Schritt nicht enthalten.

Der Prozess der Modellfindung ist also eine Mischung aus dem automatischen Vorgehen mittels der Stepwise Regression und manueller Auswahl der Prädiktoren in Hinblick auf eine geringe Anzahl an erklärenden Größen, eine einheitliche Spezifikationen in allen Konfigurationen und ein ausreichend hohes Modellgütemaß. Dabei stellt sich

zwar heraus, dass aufgrund der geringeren Anzahl an Prädiktoren ein nicht so gutes Modellgütemaß wie durch die Stepwise Regression erreicht wird. Im Hinblick auf deren Vergleichbarkeit, Einfachheit und geringere Fehleranfälligkeit werden diese Modelle dennoch bevorzugt.

#### 4.6.2 Sonderfall AbgNOx

Ein besonderer Fall ist die Zielgröße AbgNOx. Da Modelle mit den Messgrößen im Originalzustand keine zufriedenstellenden Ergebnisse lieferten und auch schon anhand der Scatterplots in Abschnitt 3.2 nichtlineare Zusammenhänge zu sehen waren, werden hier Transformationen der Prädiktoren angewandt. Welche Transformation dabei anzuwenden ist, wird mit Hilfe der Scatterplots gut ersichtlich:

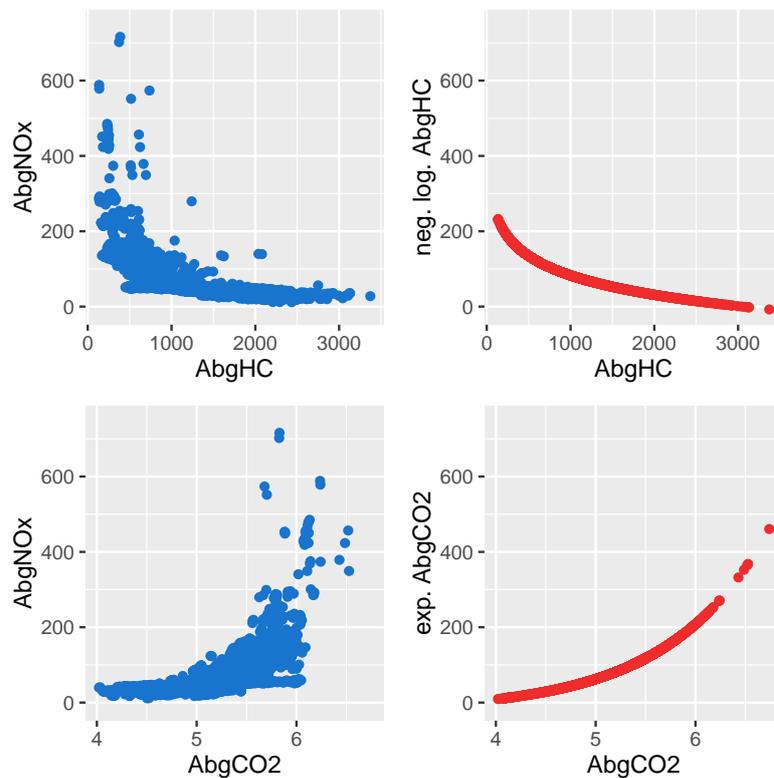


Abbildung 4.2: Analyse von Scatterplots zur Modellbildung für die Zielgröße AbgNOx

Die Messgröße AbgNOx steht in einem exponentiellen Zusammenhang mit der Messgröße AbgCO2. Der Zusammenhang mit der Messgröße AbgHC kann hingegen über einen (negativen) Logarithmus gut beschrieben werden. Um die Verbesserung im Modell zu verdeutlichen, die sich durch derartige Transformationen ergibt, folgt die Ausgabe des adjustierten Bestimmtheitsmaßes bei Aufruf des linearen Modells (`lm`) und der Modellzusammenfassung (`summary`), die unter anderem dieses enthält, in R:

## 4 Modelle

```
> model1 <- lm(AbgNOx ~ AbgHC * AbgCO2)
> model2 <- lm(AbgNOx ~ log(AbgHC) * exp(AbgCO2))
> summary(model1)$adj.r.squared
  0.7164176
> summary(model2)$adj.r.squared
  0.778338
```

Das signifikant höhere adjustierte Bestimmtheitsmaß bestätigt die Wahl der Transformationen beider Prädiktoren. In das Modell zu `AbgNOx` wird weiters die Messgröße `aGasVK` hinzugenommen.

### 4.6.3 Konfigurationenanalyse

Bezüglich der Konfigurationen ergeben die Analysen, dass es einen deutlichen Unterschied in den Messreihen bzgl. Kolben und Nockenwelle gibt. Andere Motorbauteile wie etwa das Gasventil oder die Vorkammer zeigen in keinem der Modelle einen deutlichen Unterschied auf. Wie in Abschnitt 4.6.1 gezeigt, wurde der Unterschied in den beiden Bauteilen bereits bei der Modellfindung berücksichtigt. Die Verwendung von Modellen mit Indikatorvariablen soll nun noch einmal die Notwendigkeit dieser Unterscheidung verdeutlichen. Außerdem könnten damit die Unterschiede zwischen den Konfigurationen über die Schätzungen auch qualitativ bewertet werden, was in diesem Fall jedoch nicht das vorrangige Ziel ist.

Wie in Abschnitt 2.4 beschrieben, gibt es zehn Kombinationen der beiden Bauteile Nockenwelle und Kolben. Da somit nicht jede mögliche Kombination vorliegt, ist eine Formulierung mit eigenen Regressoren für jedes Bauteil und einer eingebundenen Wechselwirkung nicht möglich. In einem Modell, das diese Bauteile als qualitative Charakteristika über Indikatorvariablen berücksichtigt, kann man aber 9 dummy-Regressoren verwendet. Die Wechselwirkung ist dabei bereits über die Indikatorvariablen selbst mitaufgenommen.

Diese Unterschiede können, wie in Abschnitt 4.4 erläutert, grafisch über Residuenplots dargestellt werden. Während bei den Modellen ohne Indikatorvariablen eine deutliche Struktur zu sehen sein sollte, müsste sich diese durch die in den neuen Modellen inkludierten *dummys* aufheben. Mit deren Hilfe werden nämlich die unterschiedlichen Schätzungen in den einzelnen Konfigurationen berücksichtigt. Im Gegensatz dazu wird beim einheitlichen, mit allen Datenpunkten geschätzten Modell über alle Konfigurationen gemittelt. Um die Verbesserung, sprich die deutlich unauffälligere Struktur in den Residuen, darzustellen, folgen nun die Residuenplots zu den Modellen ohne und mit

den 9 Indikatorvariablen für die Nockenwelle-Kolben-Kombinationen.

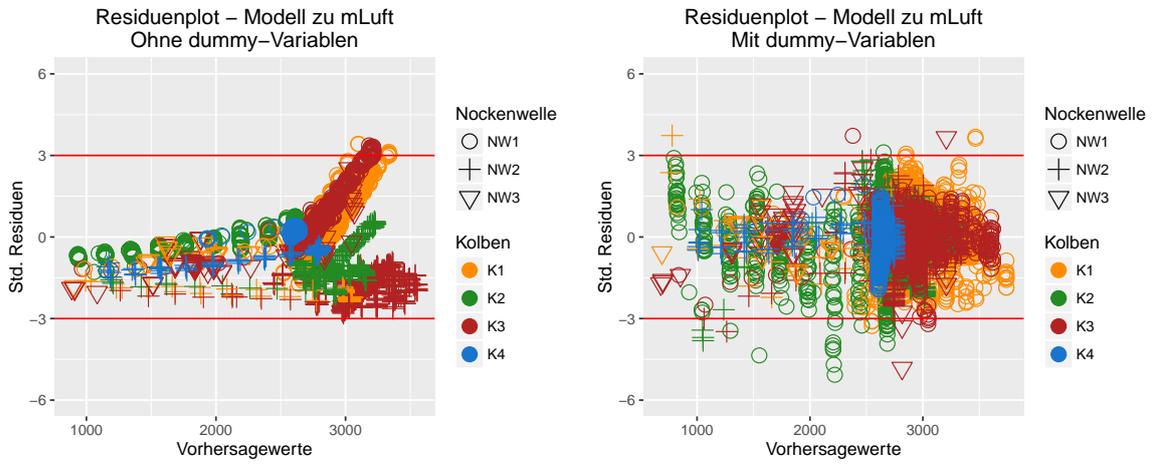


Abbildung 4.3: Scatterplots zur Konfigurationenanalyse von mLuft

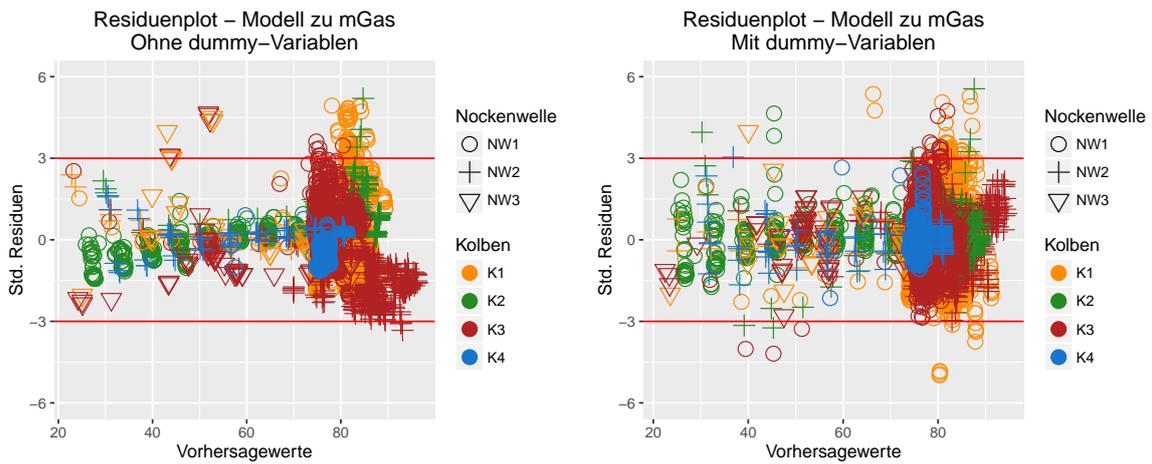


Abbildung 4.4: Scatterplots zur Konfigurationenanalyse von mGas

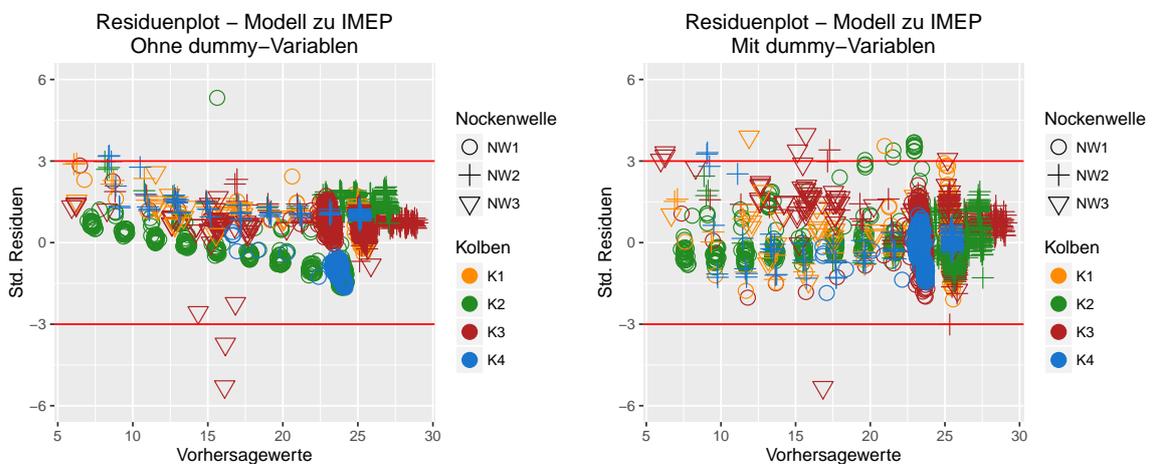


Abbildung 4.5: Scatterplots zur Konfigurationenanalyse von IMEP

Durch die Verwendung der Indikatorvariablen ist das Muster im Residuenplot deutlich zufälliger und es ist kaum noch Struktur zu erkennen. Dies ist besonders bei den drei Messgrößen  $m_{Luft}$ ,  $m_{Gas}$  und  $IMEP$  der Fall. Für die anderen Zielgrößen sehen die Ergebnisse ähnlich zu denen von  $AbgCO$  aus.

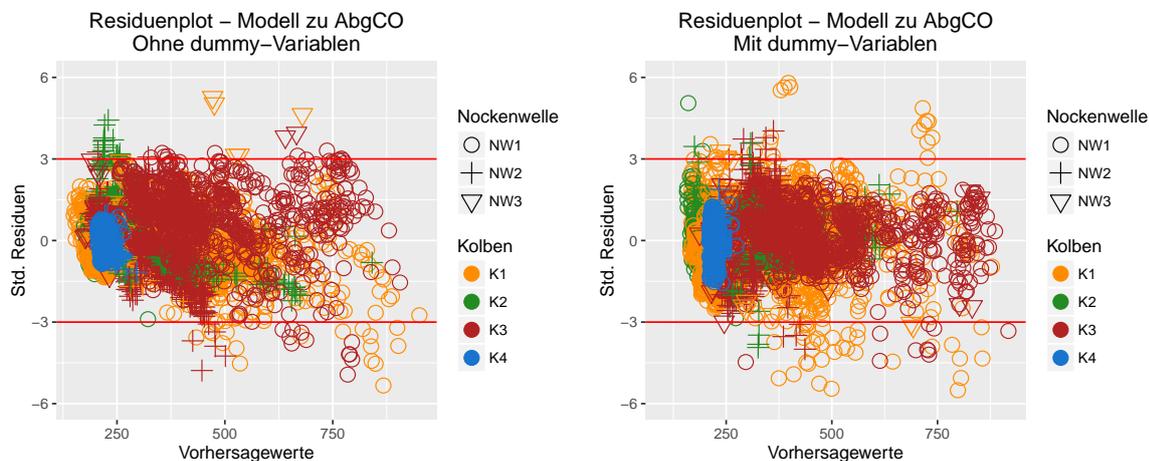


Abbildung 4.6: Scatterplots zur Konfigurationenanalyse von  $AbgCO$

Hier sieht man kaum eine Veränderung in der Residuenstruktur durch Aufnahme der Indikatorvariablen ins Modell.

#### 4.6.4 Finale Modelle

Die folgende Tabelle zeigt nun welche Modelle für die 9 Zielgrößen gewählt werden (: bezeichnet die Wechselwirkung zweier Prädiktoren), die Schätzungen der Parameter, der Standardfehler der Regression (SE) und das adjustierte Bestimmtheitsmaß ( $R_{adj}^2$ ), wenn alle 7945 Messpunkte die Basis der Modellschätzung bilden. Bei den Messgrößen  $m_{Luft}$ ,  $m_{Gas}$  und  $IMEP$  wird zusätzlich noch die Verbesserung der Modelle in SE und  $R_{adj}^2$  durch Hinzunahme der Indikatorvariablen dargestellt.

Während die Modelle für  $m_{Luft}$ ,  $m_{Gas}$  und  $IMEP$  in ihrer Form eher einfach sind, sind die Modelle der Abgaskomponenten komplexer und enthalten auch Wechselwirkungen. Weiters war es nicht möglich die Modelle der Zielgrößen  $AbgCO$  und  $AbgNO_x$  ohne die anderen Zielgrößen als erklärende Variablen zu beschreiben. Beim  $AbgNO_x$  wurde weiters eine Transformation der Prädiktoren angewandt. Wie im Abschnitt 3.2 zu sehen war, gibt es einen starken linearen Zusammenhang der Abgaskomponenten  $AbgCO_2$ ,  $AbgO_2$  und  $\lambda$ , daher sind die Modelle in ihrer Spezifikation auch dieselben und die 3 Zielgrößen können über dasselbe Modell beschrieben werden.

mLuft		mGas		IMEP	
Intercept	859.72	Intercept	-1.05	Intercept	1.82
p2	230.41	M	0.01	M	0.005
M	0.21	p3	3.38		
T2	-5.81	T3	0.03		
		p2	-1.21		
		T2	-0.02		
SE	124.64	SE	0.95	SE	0.36
$R_{adj}^2$	0.88	$R_{adj}^2$	0.99	$R_{adj}^2$	0.98
SE (Indikator)	16.36	SE (Indikator)	0.38	SE (Indikator)	0.15
$R_{adj}^2$ (Indikator)	0.99	$R_{adj}^2$ (Indikator)	0.99	$R_{adj}^2$ (Indikator)	0.99
AbgO2		Lambda		AbgCO2	
Intercept	19.21	Intercept	3.46	Intercept	0.65
T3	-0.02	T3	-0.004	T3	0.01
T4	0.009	T4	0.002	T4	-0.005
p3	0.37	p3	0.06	p3	-0.22
p2	-0.18	M	-0.00004	p2	0.12
M	-0.00005	p2	-0.0004	M	0.00003
SE	0.19	SE	0.03	SE	0.11
$R_{adj}^2$	0.89	$R_{adj}^2$	0.91	$R_{adj}^2$	0.90
AbgCO		AbgNOx		AbgHC	
Intercept	-474.05	Intercept	-90.63	Intercept	5569.42
AbgHC	0.20	exp(AbgCO2)	2.64	T3	-8.52
T3	0.99	aGasVk	-9.79	M	-2.08
p3	14.51	log(AbgHC)	19.20	p2str	0.002
AbgHC:p3	0.01	log:exp	-0.36	p3	0.03
		p3		T3:M	0.003
				T3:p2	-2.88
				T3:p3	-0.67
SE	32.71	SE	23.76	SE	296.3
$R_{adj}^2$	0.93	$R_{adj}^2$	0.79	$R_{adj}^2$	0.66

Tabelle 4.2: Analyse der Modelle der 9 Zielgrößen,  $n = 7945$ 

Weiters konnte gezeigt werden, dass die Modelle für die Abgaskomponenten über alle Messreihen hinweg gültig sind, und sich die Parameterschätzer nicht deutlich unterscheiden. Bei den Modellen zu den Zielgrößen **mLuft**, **mGas** und **IMEP** hingegen ist die Unterscheidung bzgl. der Motorbauteile Nockenwelle und Kolben deutlich notwendig. Dies macht auch physikalisch Sinn, da diese Einfluss auf den Liefergrad haben, der

wiederum genau diese Messgrößen beeinflusst. Es kann weiters mit Hilfe der Konfigurationenanalyse gefolgert werden, dass ein Wechsel der Datenbasis für die Schätzung der Modellparameter in der Praxis nur dann erforderlich ist, wenn die Bauteile Nockenwelle und/oder Kolben getauscht werden.

### 4.6.5 Evaluierung

Zum Abschluss werden nun die Residuenplots zu den Modellen betrachtet, wobei diese bezüglich der Messpunkte getrennt nach Nockenwelle und Kolben gebildet werden. Hier die Ergebnisse zu den Messpunkten der Messreihen von Konfiguration 1.

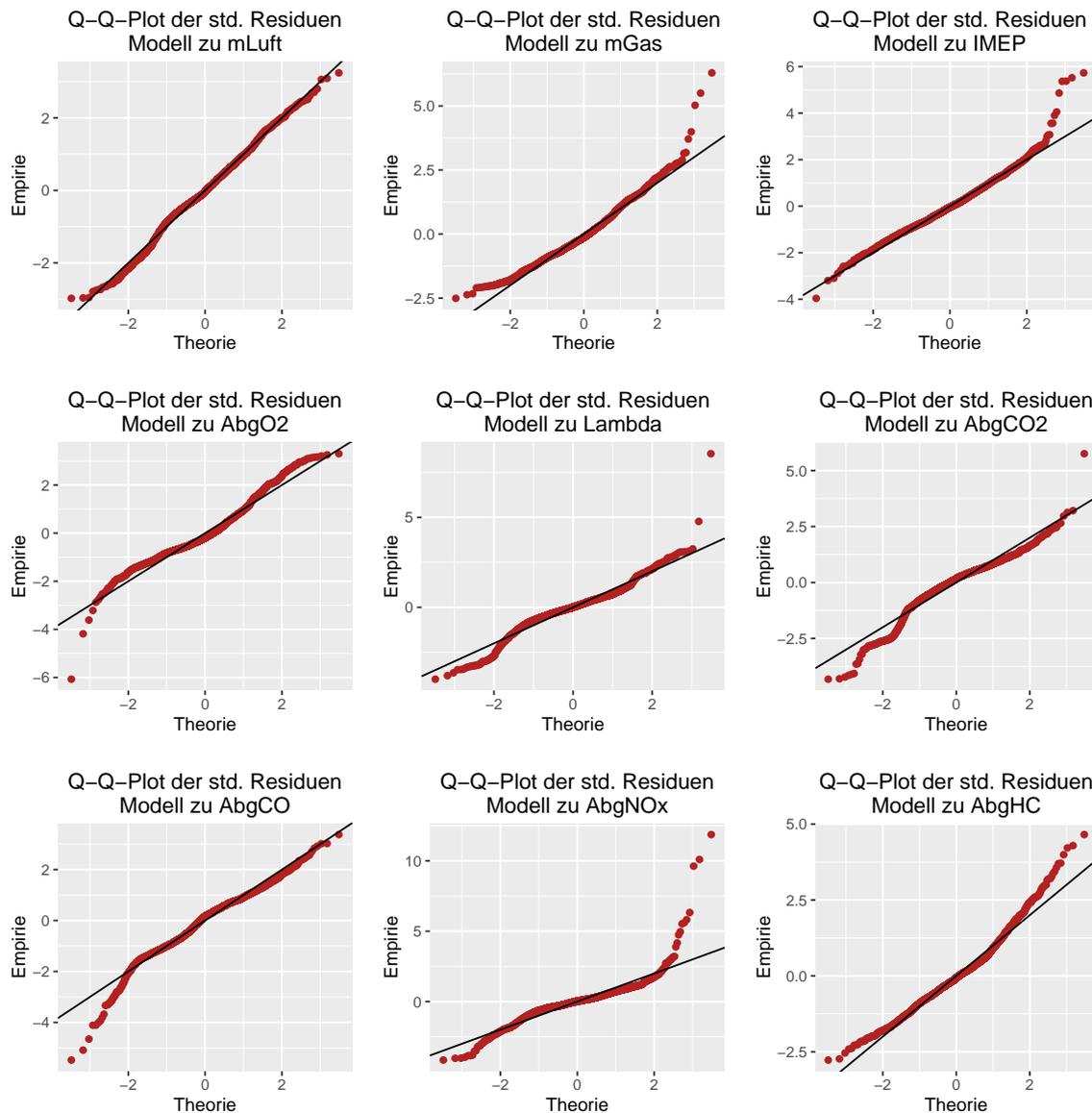


Abbildung 4.7: Q-Q-Plots der standardisierten Residuen

In den meisten Fällen sind die standardisierten Residuen zwar im Zentrum symmetrisch verteilt, die Normalverteilungsannahme kann aber aufgrund der Schwänzigkeit und der Ausreißer nicht bestätigt werden. Diese wird vor allem im Zuge von Hypothesentests bzgl. der Modelle verwendet, was jedoch nicht Ziel dieser Arbeit ist. Bei den Messpunkten der anderen Konfigurationen erhält man ein ähnliches Bild.

Die Plots der Residuen gegen die Vorhersagewerte zur Überprüfung der Unabhängigkeit der beiden Größen liefern folgende Ergebnisse:

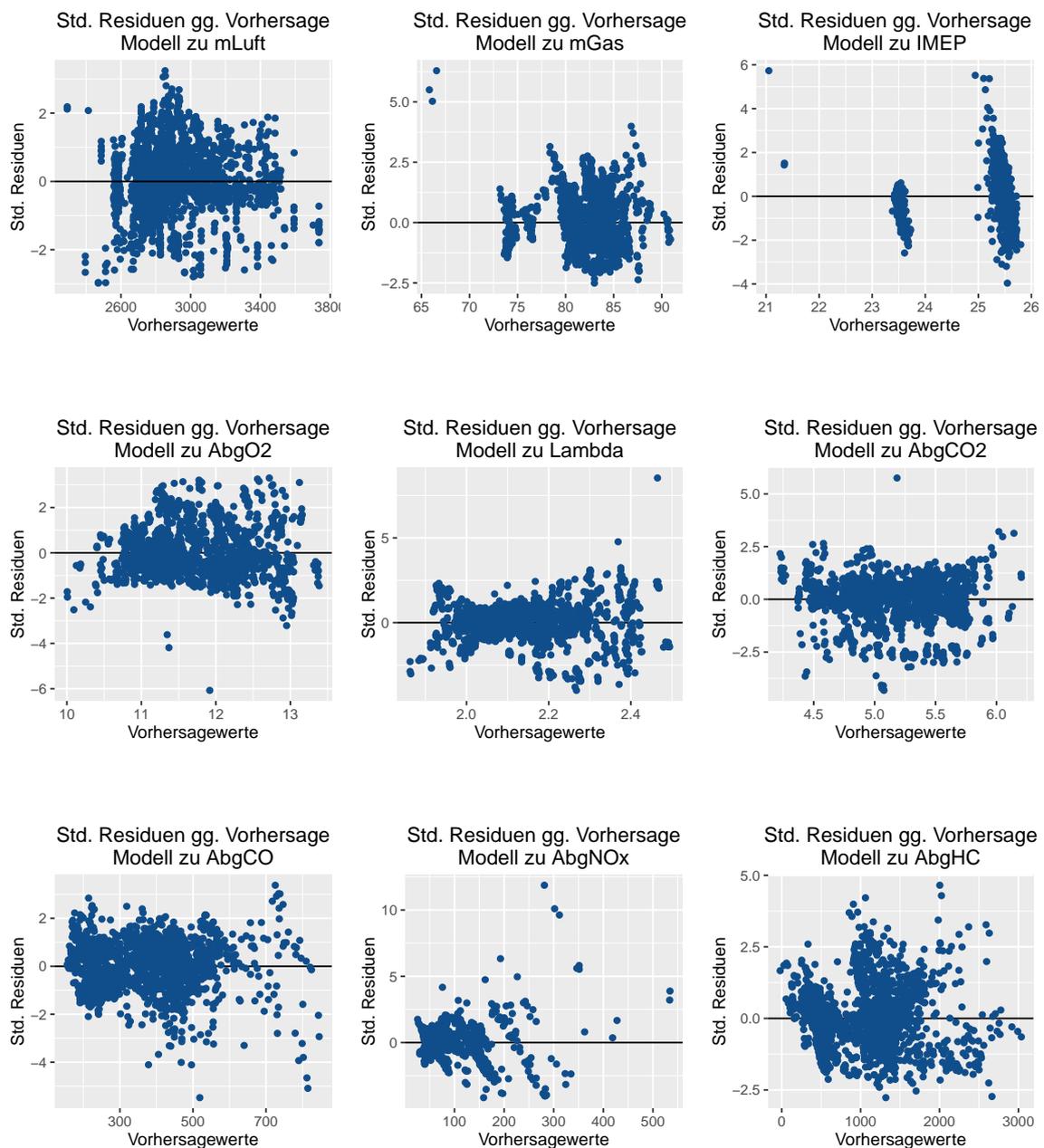


Abbildung 4.8: Scatter-Plots: Stand. Residuen gegen Vorhersagewerte

Bis auf einige Ausreißer in manchen Plots ist kaum eine Struktur zu erkennen. Lediglich bei der Messgrößen **IMEP** können 2 Gruppen von Punkten identifiziert werden. Dies liegt aber daran, dass dies eine Messgröße ist, die meist auf unterschiedliche Stufen hingeregelt wird, wie in Abschnitt 3.1 gezeigt. Die Messgröße **AbgNOx** zeigt die meisten Ausreißer. In allen Fällen scheint eine Transformation der Zielgröße nicht notwendig.

# 5 Messdatendiagnose

Das Hauptaugenmerk dieser Arbeit liegt darauf eine Messdatendiagnose zu entwickeln, die es ermöglicht die Messdaten auf Basis bestimmter Kriterien in zwei Klassen einzuteilen - fehlerhaft oder fehlerfrei. Die Idee der hier angewandten Methode besteht darin, Prognosen für zukünftige Werte zu machen und diese mit den tatsächlichen Werten zu vergleichen. Die Prognosen können dabei mit den im vorherigen Kapitel vorgestellten Modellen gemacht werden. Wie diese und deren Fehler genau aussehen wird in Abschnitt 5.1 erläutert. Zur Diagnose an sich wird dann nur der Fehler bzw. der relative Fehler der Prognose herangezogen. Für die Klassifizierung eines Messpunktes wurden 3 unterschiedliche Kriterien entwickelt. Diese sind eine Kombination aus der Theorie der Kontrollkarten und verschiedener Prognosegütemaße. Zum Abschluss werden Details zur Umsetzung in der Praxis gezeigt.

## 5.1 Prognose und Prognosefehler

Aufgrund der Daten zu den vergangenen Messungen der Prädiktorvektoren  $x_1, \dots, x_{t-1}$  und auch der Zielgröße  $y_1, \dots, y_{t-1}$  werden die Parameter zu den bereits spezifizierten, linearen Regressionsmodellen über die Kleinste-Quadrate-Methode geschätzt. Ziel ist nun für den Punkt  $x_t$  eine Vorhersage, also eine Prognose von  $y_t$ , zu machen. Man spricht hier laut HACKL (2005) von einer sogenannten 1-Schritt-Prognose. Wenn man annimmt, dass das Modell auch im Punkt  $t$  gültig ist

$$y_t = x_t' \beta + \epsilon_t \quad \text{mit} \quad \epsilon_t \sim N(0, \sigma^2)$$

und die Prädiktoren  $x_t$  und der Schätzer  $\hat{\beta}$  für  $\beta$  bekannt sind, können die Prognosewerte ermittelt werden über

$$\hat{y}_t = \hat{\mu}_t = x_t' \hat{\beta}.$$

Der Prognosewert entspricht also dem geschätzten Erwartungswert.

Nimmt man an, dass die Modellannahmen und die Folgerungen daraus auch für den Prognosezeitraum gültig sind und dass weiters die neue Beobachtung  $y_t$  unabhängig

von den abhängigen Größen  $y_1, \dots, y_{t-1}$  ist, dann kann gefolgert werden, dass für die Prognose bzw. den Prognosefehler gilt

$$\begin{aligned}\mathbb{E}(y_t - \hat{y}_t) &= \mathbb{E}(y_t) - \mathbb{E}(\hat{y}_t) = x_t' \beta - x_t' \beta = 0, \\ \text{Var}(\hat{y}_t - y_t) &= \text{Var}(y_t) + \text{Var}(\hat{y}_t) = \sigma^2 + \sigma^2 (x_t (X'X)^{-1} x_t') = \sigma^2 (I_n + x_t (X'X)^{-1} x_t').\end{aligned}$$

Dies folgt vor allem aus den in Abschnitt 4.2 gezeigten Resultaten der Kleinste-Quadrate-Schätzung und kann auch in SEN et al. (1990) nachgelesen werden. Damit ist die Verteilung, genauer der Erwartungswert und die Varianz, des Prognosefehlers bekannt.

Im Anschluss an die Schätzung der Parameter des Modells und der daraus resultierenden Vorhersage, kann diese mit dem tatsächlich gemessenen Wert, so bald dieser vorhanden ist, „verglichen“ werden. Man erhält dann den tatsächlichen Prognosefehler. Für die 1-Schritt-Vorhersagen definieren wir den Prognosefehler als eindimensionalen Wert gegeben durch

$$e_t = y_t - \hat{y}_t.$$

Dies bezeichnet also die absolute Abweichung von der Vorhersage zum tatsächlich gemessenen Wert.

Der absolute Fehler der Vorhersage hängt von der Skalierung der Beobachtungen  $y$  ab. Da im Zuge des Diagnoseverfahrens jedoch mehrere unterschiedlich skalierte Größen klassifiziert werden sollen, ist es sinnvoll eine modifizierte Definition des Fehlers anzuwenden. Dies kann etwa über die Betrachtung des relativen Vorhersagefehlers passieren, der unter anderem berechnet werden kann, in dem der absolute Vorhersagefehler durch den tatsächlich beobachteten Wert dividiert wird:

$$re_t = \frac{y_t - \hat{y}_t}{y_t}.$$

Die Abweichung von der Vorhersage zum tatsächlichen Messwert wird also in Relation zum beobachteten Wert gesetzt. In der Literatur gibt es noch andere Formen der Relativierung, etwa in dem durch den Vorhersagewert dividiert wird.

## 5.2 Prognosegütemaße

In Analogie zu einem Konfidenzintervall für einen Punktschätzer kann das Prognoseintervall für die Prognosewerte bzw. das Prognoseintervall der Fehler für den Fehler der Prognose verwendet werden. Man kann also mit Hilfe der Verteilungseigenschaften einen sogenannten Vertrauensbereich für den Fehler definieren:

$$0 \pm \text{factor } \sigma \left( I_n + x_t(X'X)^{-1}x_t' \right).$$

Kann von einer Normalverteilung der Prognosefehler ausgegangen werden, dann ist der Faktor gleich dem Normalverteilungsquantil zum jeweils gewählten Niveau  $\alpha$ . Laut FRIEDL (2005) ist ein solches Vorhersageintervall dann ein Intervall, welches mit Wahrscheinlichkeit  $(1 - \alpha)$  die Realisierung dieser zusätzlichen Response beinhalten wird.

In der Praxis wird weiters ein Schätzer für die unbekannte Varianz  $\sigma^2$  bzw. für die Standardabweichung  $\sigma$  benötigt. Für den absoluten Fehler wurde dieser bereits in Abschnitt 4.2 hergeleitet. Dort wurde gezeigt, dass die modifizierte Version der Residuenquadratsumme,  $SSE(\hat{\beta})/(n - k)$ , ein erwartungstreuer Schätzer von  $\sigma^2$  ist. Für die Standardabweichung wird diese Schätzung auch als „Root Mean Squared Error“ bezeichnet und ist definiert als:

$$RMSE = \sqrt{\frac{1}{n - p} \sum_{i=1}^n (\hat{\mu}_i - y_i)^2}.$$

Für den relativen Fehler kann dieser folgendermaßen zum „Relative Root Mean Squared Error“ modifiziert werden

$$RRMSE = \sqrt{\frac{1}{n - p} \sum_{i=1}^n \left( \frac{\hat{\mu}_i - y_i}{y_i} \right)^2},$$

wobei  $n$  die Anzahl aller Messpunkte, auf die das aktuelle Modell basiert, und  $p$  die Anzahl der zu schätzenden Parameter im Modell beschreibt.  $y_i$  definiert den tatsächlichen Wert an Datenpunkt  $i$  und  $\hat{\mu}_i$  die Schätzung bzw. die Vorhersage an diesem Messpunkt.

Neben diesem Vertrauensbereich bzw. Prognoseintervall sind zur Beurteilung der Prognosequalität in der Literatur auch noch andere Maßzahlen bekannt. Diesen widmen sich etwa die Werke MARÉCAL et al. (2015) und ALBERS et al. (Hrsg.) (2006). In der folgenden Tabelle sind einige Beispiele solcher Prognosegütemaße samt Abkürzungen und Formeln dargestellt. Dabei beschreibt  $n$  die Anzahl der Vorhersagen und  $o_i$  die beobachteten bzw.  $f_i$  die vorhergesagten Messwerte im beobachteten Zeitraum.

Maß	Abk.	Formel	Bereich
Mean (Absolut Gross) Error	ME	$\frac{1}{n} \sum_{i=1}^n  f_i - o_i $	$[0, \infty]$
Mean Normalized (Gross) Error	MNE	$\frac{1}{n} \sum_{i=1}^n \frac{ f_i - o_i }{o_i}$	$[0, \infty]$
Normalized Mean Error	NME	$\frac{\sum_{i=1}^n  f_i - o_i }{\sum_{i=1}^n o_i}$	$[0, \infty]$
Mean Fractional Error	MFE	$\frac{2}{n} \sum_{i=1}^n \left  \frac{f_i - o_i}{f_i + o_i} \right $	$[0, 2]$
Mean Bias	MB	$\frac{1}{n} \sum_{i=1}^n (f_i - o_i)$	$[-\bar{o}, \infty]$
Mean Normalized Bias	MNB	$\frac{1}{n} \sum_{i=1}^n \frac{(f_i - o_i)}{o_i}$	$[-1, \infty]$
Normalized Mean Bias	NMB	$\frac{\sum_{i=1}^n (f_i - o_i)}{\sum_{i=1}^n o_i}$	$[-1, \infty]$
Mean Fractional Bias	MFB	$\frac{2}{n} \sum_{i=1}^n \frac{(f_i - o_i)}{(f_i + o_i)}$	$[-2, 2]$

Tabelle 5.1: Prognosegütemaße

Der Mean Error und der Mean Bias sind Mittelwerte der (absoluten) Fehler. Beim Mean Bias heben sich negative und positive Abweichungen auf. Er kann als Maß für die Verzerrtheit der Vorhersagen gesehen werden. Werte um Null indizieren, dass die Vorhersagen unverzerrt sind. Hohe positive bzw. negative Werte hingegen sind Hinweis für eine Über- bzw. Unterschätzung der Werte durch die Prognose. Der Mean Error beschreibt wie groß im Mittel die absoluten Abweichungen sind. Negative und positive Werte heben sich hier nicht auf. Dieses Maß ist somit eher dafür geeignet, um Aussagen darüber zu treffen, wie sehr die Prognosen vom tatsächlichen Wert abweichen. Wie auch der rohe Prognosefehler sind diese beiden Maße von der Skalierung der beobachteten Werte abhängig.

Daher ist es wiederum sinnvoll Maße bezogen auf die relativen Fehler zu definieren, die sowohl der MNE als auch der MNB liefert. Diese beiden Maße bieten also auch den Vorteil, dass sie dimensionslos und unabhängig von der Größenordnung der betrachteten Werte sind. Der Mean Normalized Error ist auf Grund seiner Darstellung gleich dem Mittelwert der Absolutbeträge der relativen Fehler. Der Mean Normalized Bias ist der Mittelwert der tatsächlichen relativen Fehler. Die Darstellung beider Maße als Mittelwerte

$$\begin{aligned}
 are_i = \frac{|f_i - o_i|}{o_i} &\Rightarrow \text{MNE} = \overline{are} = \frac{1}{n} \sum_{i=1}^n are_i \\
 re_i = \frac{f_i - o_i}{o_i} &\Rightarrow \text{MNB} = \overline{re} = \frac{1}{n} \sum_{i=1}^n re_i
 \end{aligned}$$

ist vor allem für deren Verwendung in der anschließenden Fehlerdiagnose von Vorteil. Näheres dazu wird im nächsten Abschnitt erklärt. Je niedriger dabei die Realisierungen dieser beiden Maße (im Absolutbetrag) sind, desto besser ist die Qualität der Prognosen.

Der Normalized Mean Bias bzw. Error und der Mean Fractional Bias bzw. Error beschreiben noch andere Formen der Standardisierung. Der Mean Fractional Bias und Error bieten dabei den Vorteil, dass deren Realisierungen standardisiert sind, wie auch in der Tabelle ersichtlich. Dies ist vor allem beim Vergleich unterschiedlicher Diagnoseverfahren von Vorteil, was hier allerdings nicht der Fall ist.

## 5.3 Diagnoseverfahren

Ziel des Vergleiches der Vorhersage mit dem tatsächlichen Messwert ist es, die Messpunkte in 2 Klassen einteilen zu können. So sollten die Messpunkte entweder als fehlerfrei oder als fehlerhaft gemeldet werden. Das in dieser Arbeit vorgestellte Klassifizierverfahren ist an die Methoden der sogenannten Kontrollkarten angelehnt (siehe dazu etwa MONTGOMERY (2009)). Dabei werden die zu untersuchenden Größen im Zeitverlauf geplottet. Der Plot enthält außerdem eine sogenannte „center line“, die den erwartenden Mittelwert der Größe widerspiegelt und zwei weitere Linien, die die obere und untere Kontrollgrenze darstellen bzw. den bereits erwähnten Vertrauensbereich definieren. Dieses Prinzip wurde im Rahmen dieser Arbeit auf die relativen Prognosefehler angewandt. Die „center line“ entspricht der Null. Die Kontrollgrenzen werden über den im vorherigen Abschnitt hergeleiteten Vertrauensbereich des Prognosefehlers definiert.

Mit Hilfe der Kontrollkarten sollen nun sogenannte „out-of-control“ Bedingungen hergeleitet werden. In diesem Fall bedeutet dies, dass die Annahme eines fehlerhaften Messpunktes getätigt werden kann. Laut MONTGOMERY (2009) kann eine „out-of-control“ Situation etwa dann vorliegen, wenn einer oder mehrere Punkte in Folge außerhalb des Vertrauensbereiches liegen oder wenn die Punkte ein nicht zufälliges Verhalten aufweisen. Durch eine gewisse Menge an Entscheidungsregeln sollen nun solche nicht zufällige Strukturen der relativen Prognosefehler erkannt werden.

Auf diese Arbeit bezogen kann vermutlich darauf geschlossen werden, dass der Prozess außer Kontrolle bzw. fehlerhaft ist, falls eine der drei in Folge beschriebenen Kriterien beobachtet wird. Die Entwicklung dieser Kriterien geschah unter anderem mittels der Analyse der relativen Prognosefehler zu Daten, in die gezielt Fehler eingebaut wurden. Die grafischen Beispiele etwa stammen aus solchen Analysen. Weiters konnte man damit beobachten, bei welchen Fehlern welche Kriterien am effizientesten sind und wie die 3 Kriterien sich zueinander verhalten. Das 1. Kriterium bezieht sich dabei auf Werte außerhalb des Bereiches. Das 2. und 3. Kriterium sind eher darauf abgezielt Muster in den relativen Prognosefehlern aufzudecken.

### 1. Kriterium

Situation: Es liegen mehrere relative Fehler in Folge außerhalb der Grenze.

Dies kann etwa bei einem mittleren bis sehr starken Sprung in den Messdaten passieren. Die Erkennung ist hierbei recht einfach. So kann diese direkt durch Betrachtung des relativen Fehlers und des Vertrauensbereiches geschehen. Die Besonderheit ist dabei, dass mehrere, unterschiedlich große Vertrauensbereiche definiert werden, um dieses Kriterium so sensibel wie möglich zu gestalten. Dabei gilt natürlich, dass je kleiner der Bereich ist, desto größer ist die Anzahl der Werte die außerhalb liegen dürfen, ohne den Verdacht eines Fehlers im System zu erheben.

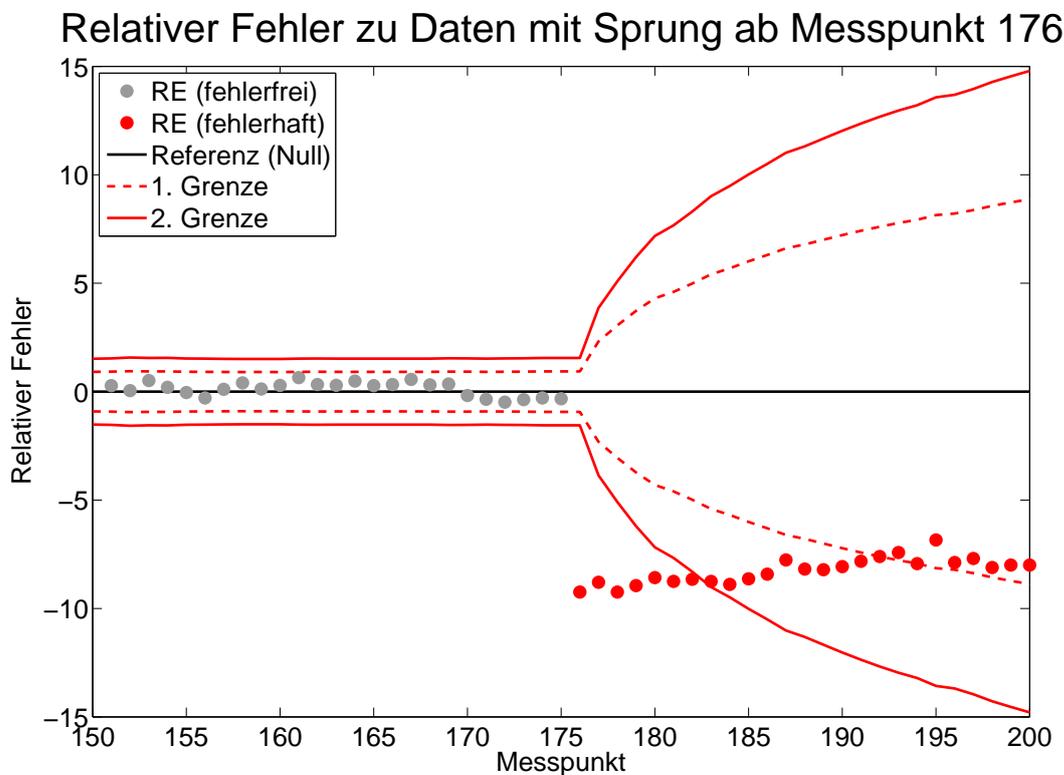


Abbildung 5.1: Kriterium 1 - Großer Vertrauensbereich

Anhand dieser Grafik wird ersichtlich, dass sehr große Fehler, so weit sie nicht erkannt und deshalb in die Modellbildung miteinbezogen werden, eine Vergrößerung des Vertrauensbereiches mit sich ziehen, da auch die Residuenquadratsumme des Modells mit den fehlerhaften Messpunkten ansteigt. Deswegen muss die Anzahl der Punkte die in Folge außerhalb liegen dürfen, so gering wie möglich eingestellt werden.

Es wird ein kleiner und ein großer Bereich eingeführt, um etwaige Fehler so schnell wie möglich zu erkennen. Der große Bereich ist notwendig, um sehr große Fehler unmittelbar zu erkennen. Würde es nur den kleinen Bereich geben, so würden sehr große Fehler, wegen der höheren Anzahl erlaubter aufeinanderfolgender Werte außerhalb, zu spät erkannt. Diese würden in die Modellbildung aufgenommen werden, was zu einem starken Sinken des Modellgütemaßes führen würde. Damit könnten keine vernünftigen Vorhersagen mehr gemacht werden. Dies würde auch zu einem extremen Anwachsen des Vertrauensbereiches führen, da die Modelle auch mit den fehlerhaften Werten gebildet werden, sodass wiederum keine neuen Fehler erkannt werden könnten. Hätte man hingegen nur den großen Bereich, so könnte man in Folge nur sehr große Fehler aufdecken. Situationen, wie in der nächsten Grafik dargestellt, könnten nicht als fehlerhaft klassifiziert werden.

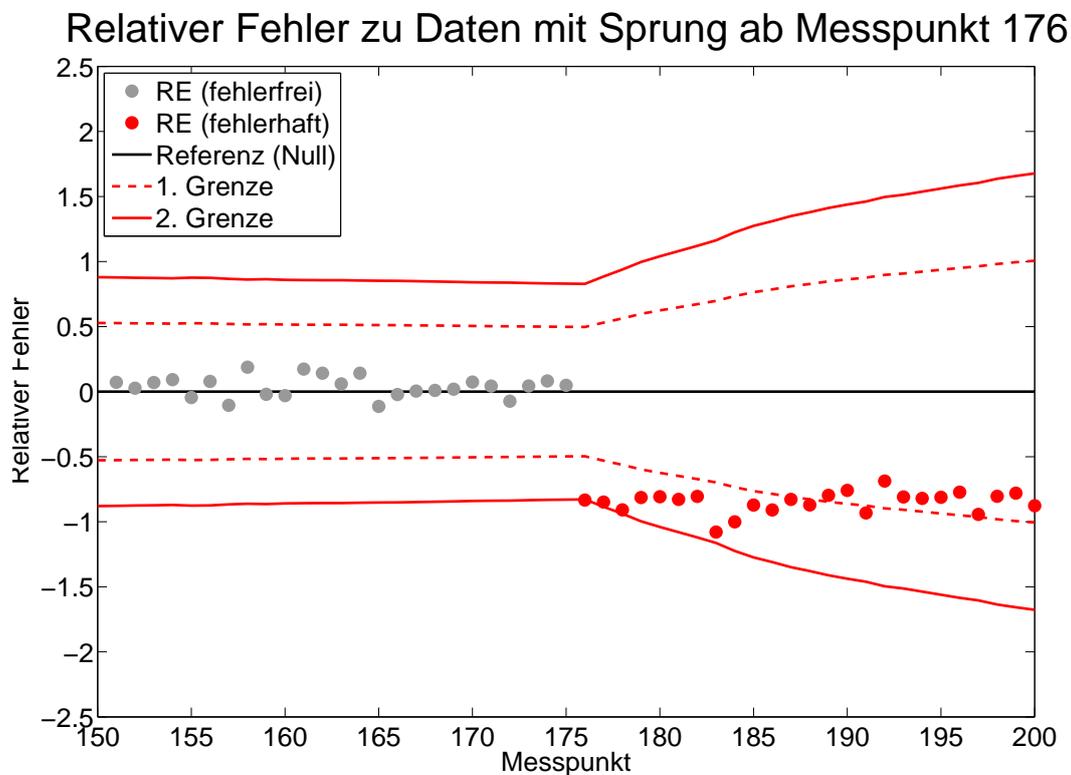


Abbildung 5.2: Kriterium 1 - Kleiner Vertrauensbereich

In der Praxis wird außerdem noch ein dritter Vertrauensbereich verwendet, dessen Grenzen jedoch extrem weit eingestellt sind. Dieser ist dazu da um sehr extreme Ausreißer sofort zu erkennen und zu eliminieren. Bereits ein Punkt außerhalb dieses Bereiches wird als fehlerhaft gemeldet.

## 2. Kriterium

Situation: Der Absolutbetrag des relativen Fehlers steigt über mehrere Messpunkte hindurch an.

Ein derartiges Szenario kann etwa bei einem stark ausgeprägten, schleichenden Fehler vorliegen. Zur Erkennung werden nicht die Werte der relativen Fehler betrachtet, sondern die Mittelwerte ihrer Absolutbeträge. Im Sinne der Prognosegütemaße entspricht das daher der Berechnung des MNE, wobei dieser nicht für alle vergangenen Prognosen sondern nur für eine gewisse Anzahl davon, sprich einer Gruppe, berechnet wird. Man möchte also eine Aussage darüber tätigen, ob der relative Fehler der Gruppe hoch ist oder nicht. Um dies in einem vernünftigen Rahmen tun zu können, wird eine gewisse Referenz benötigt.

Dies ist vor allem deswegen notwendig, da im Zuge des Diagnoseverfahrens unterschiedliche Messgrößen betrachtet werden, deren relative Fehler und somit auch der Mittelwerte sich stark unterscheiden. Betrachtet man etwa die Verläufe der relativen Fehler für die beiden Zielgrößen  $m_{\text{Gas}}$  und  $\text{AbgCO}$  so wird dieser Unterschied besonders deutlich.

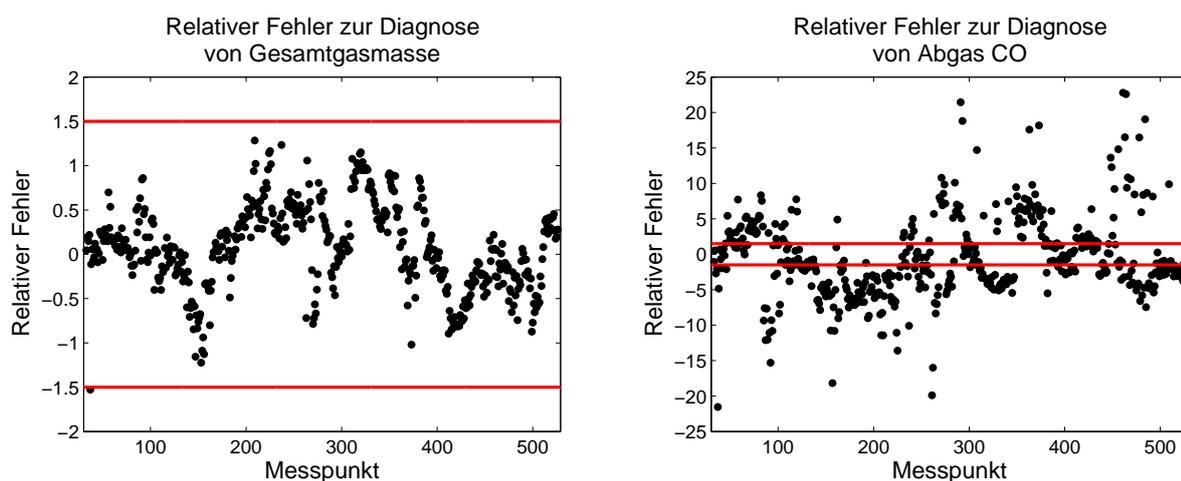


Abbildung 5.3: Unterschied im Wertebereich der relativen Fehler

Während bei der Zielgröße  $m_{\text{Gas}}$  die relativen Fehler im Bereich  $[-1.5, 1.5]$  liegen, gibt es bei der Messgröße  $\text{AbgCO}$  auch relative Fehler größer als 15. Der Bereich der relativen Fehler von  $m_{\text{Gas}}$  ist in beiden Grafiken rot markiert und verdeutlicht diesen Unterschied.

Als Referenz wird nun der Mittelwert einer Gruppe genommen, die viel größer ist als die zu analysierende Gruppe und die zu analysierende Gruppe aber enthält. Dieser wird in Folge als Gesamtmittelwert bezeichnet. Mit Hilfe dieses Wertes kann nun ein Konfidenzintervall für den Gruppenmittelwert definiert werden. Weiters wird dazu eine Schätzung der Varianz des Gruppenmittelwertes benötigt.

Da dieser, wie im vorhergehenden Abschnitt beschrieben, als Mittelwert der absoluten relativen Fehler formuliert werden kann, kann dessen Varianz über die Varianz eines Mittelwertes hergeleitet werden. Für diesen gilt nämlich

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \Rightarrow \quad \mathbb{E}(\bar{x}) = \mathbb{E}(x) \quad \text{und} \quad \text{Var}(\bar{x}) = \frac{\text{Var}(x)}{n}.$$

Bezeichne nun  $are = abs(re)$ . So gilt für den MNE

$$\mathbb{E}(MNE) = \mathbb{E}(are) = 0 \quad \text{und} \quad \text{Var}(MNE) = \frac{\text{Var}(are)}{n} = \frac{\sigma_{are}^2}{n}$$

mit  $\hat{\sigma}_{are}^2 = \frac{1}{n} \sum_{i=1}^n (are_i - \overline{are})^2$  als Schätzer für  $\sigma_{are}^2$ .

Mit Hilfe dieser Verteilungseigenschaften kann nun ähnlich dem Vertrauensbereich für den relativen Fehler ein Konfidenzintervall für den Gruppenmittelwert formuliert werden über

$$MNE_{Gesamt} + \mathbf{factor} \hat{\sigma}_{are}.$$

Man erkennt, dass es sich hierbei um einen einseitig abgegrenzten Bereich handelt, da nur besonders große und nicht sehr kleine relative Fehler erkannt werden sollen. Hier kann wiederum der **factor** über ein Normalverteilungsquantil gewählt werden, wenn man an einem bestimmten Niveau interessiert ist.

Falls der Gruppenmittelwert nicht in diesem Konfidenzintervall realisiert, kann darauf geschlossen werden, dass die Gruppe signifikant unterschiedlich zur Gesamtmenge ist. Von Bedeutung sind dabei aber nur Gruppen, die oberhalb der Grenze liegen, da das Kriterium auf die Erkennung von einem Anstieg im relativen Fehler abzielt. Wird dies für eine gewisse Anzahl von Gruppen in Folge getan, so ist dies ein Zeichen dafür, dass es im System einen Fehler gibt. Die folgende Abbildung zeigt einen Fall, für den dies zutrifft. Außerdem wird dabei das Prinzip der Gruppenbildung erkenntlich, die auch beim 3. Kriterium angewandt wird.

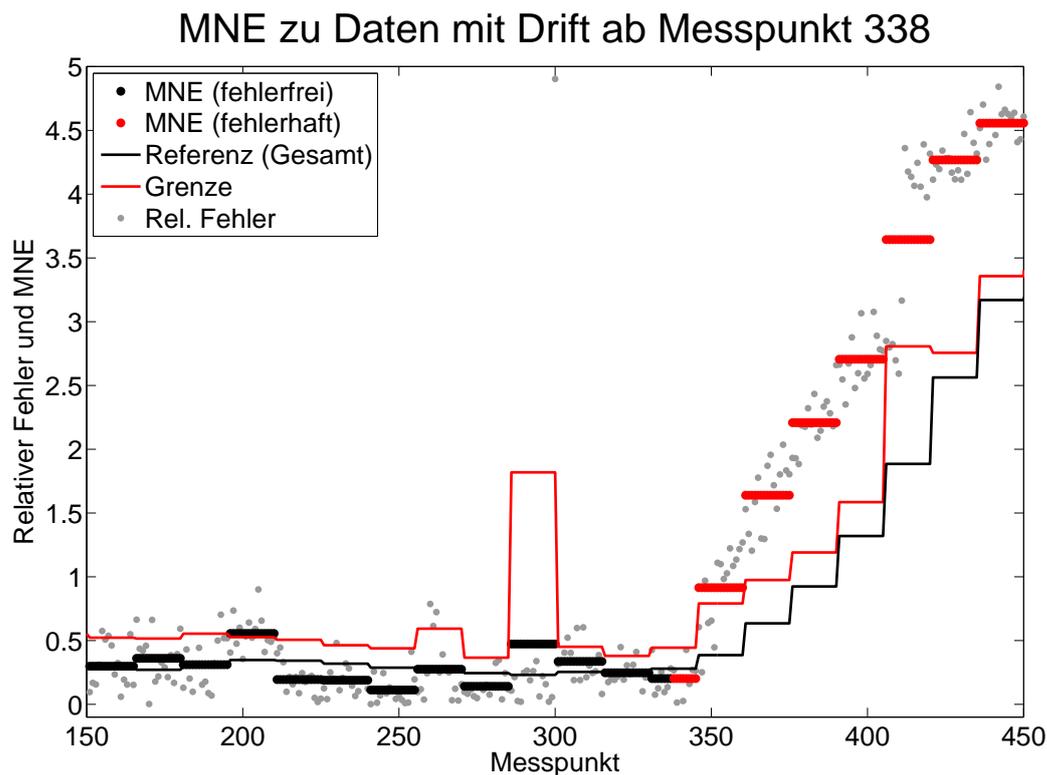


Abbildung 5.4: Kriterium 2

### **3. Kriterium**

Situation: Sehr viele Punkte hintereinander tendieren in eine Richtung, möglicherweise auch innerhalb der Grenze (langanhaltende Über- bzw. Unterschätzung).

Dies kann sowohl bei schleichenden Fehlern als auch bei Sprüngen von geringer Stärke passieren. Ähnlich der Identifikation im 2. Kriterium werden hier nicht die relativen Werte sondern deren Mittelwerte betrachtet. In diesem Fall jedoch wird über die rohen Werte gemittelt, somit entspricht dies der Berechnung des MNB. Auch hier wird dieser gruppenweise berechnet. Als Referenz dient dabei, da man auf Unverzerrtheit der Prognosen der Gruppe testen möchte, die Null. Weiters kann auch die Varianz des MNB ähnlich zur Varianz des MNE formuliert werden, da es sich auch bei diesem Maß, wie bereits erwähnt, um einen Mittelwert handelt. So kann gezeigt werden:

$$\mathbb{E}(MNB) = \mathbb{E}(re) = 0 \quad \text{und} \quad \text{Var}(MNB) = \frac{\text{Var}(re)}{n} = \frac{\sigma_{re}^2}{n}$$

mit  $\hat{\sigma}_{re}^2 = \frac{1}{n} \sum_{i=1}^n (re_i - \bar{re})^2$ . Somit kann der Vertrauensbereich formuliert werden über

$$0 \pm \text{factor } \hat{\sigma}_{re}.$$

Liegt der Gruppenmittelwert nicht in diesem Intervall, so kann die Annahme der Unverzerrtheit der Gruppe nicht bestätigt werden. Wird dies für eine gewisse Anzahl von Gruppen in Folge getan, so werden die Punkte der Gruppen als fehlerhaft klassifiziert.

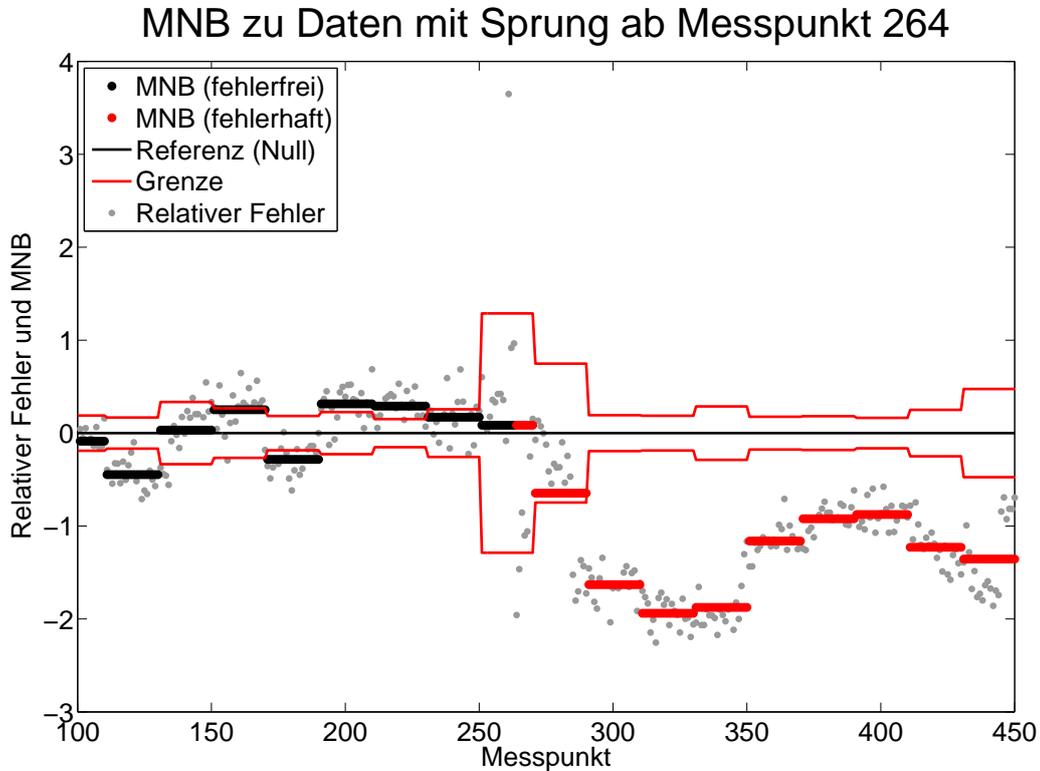


Abbildung 5.5: Kriterium 3

Man erkennt, dass der Sprung in den Daten zu einer deutlichen Unterschätzung des Modells führt, sodass mehrere Gruppenmittelwerte in Folge außerhalb des Konfidenzintervalles liegen.

## 5.4 Anmerkungen zur praktischen Umsetzung

In diesem Fall ist ein besonderes Augenmerk auf die Datenpunkte zu legen, die in die Schätzung der Parameter eingehen. Im Speziellen wird hier die Anzahl der Messpunkte, die zur Vorhersage verwendet werden, eingeschränkt. So benutzt man etwa nicht alle vergangenen Messpunkte, sondern nur eine gewisse Untermenge davon, die zeitlich näher liegenden. Dies hat den Grund, dass die Modelle in Folge mehr Bezug zu den aktuellen Daten haben sollen. Außerdem ist der Rechenaufwand etwas kleiner und dadurch kann die Laufzeit etwas gesenkt werden. Vorsicht ist dennoch geboten, da die zeitliche Bandbreite auch nicht zu gering eingestellt werden darf. Die Modelle

würden sich dann zu schnell an möglicherweise fehlerhafte Daten anpassen. Es hat sich gezeigt, dass fast sämtliche Modelle gute Resultate liefern, wenn für die Prognosewerte  $\hat{y}_t$  nur die 500 letzten vergangenen Messpunkte benutzt werden. Die Ausnahme bildet die Zielgröße IMEP. Da diese in manchen Fällen über lange Zeiträume hinweg konstant gehalten bzw. geregelt wird, werden hier alle vergangenen Messpunkte in die Modell-schätzung miteinbezogen.

Im Rahmen des Fehlerklassifizierverfahrens werden außerdem nur jene vergangenen Messpunkte in die Modellbildung einbezogen, die zum Zeitpunkt der Vorhersage als fehlerfrei gemeldet wurden. Da manche Abfragen, wie im Folgenden beschrieben, erst nach mehreren vergangenen Datenpunkten, also zeitverschoben, einen Fehler aufdecken können, kann es jedoch trotzdem passieren, dass Modelle mit fehlerhaften Messpunkten geschätzt werden.

Weiters wurde noch ein Kriterium hinzugefügt, über welches bestimmt wird, ab wann überhaupt eine Vorhersage gemacht wird. Dies sollte nur dann der Fall sein, wenn das Modell auch „einigermaßen gut“ ist. In der Praxis bedeutet dies das hier wiederum ein Modellgütemaß, genauer gesagt das adjustierte Bestimmtheitsmaß  $R^2$ , benötigt wird. Nur wenn dieses ausreichend groß ist, wird auch eine Vorhersage gemacht. Diese Schranke wurde mit 0.4 festgelegt.

Da das Fehlerklassifizierverfahren von mehreren Parametern abhängt, ist es notwendig sogenannte Referenzszenarien zu betrachten und diese Parameter durch Simulation einzustellen. Durch Betrachtung der Verläufe der relativen Fehler mit und ohne künstlich eingebaute Fehler wurde ersichtlich welche Szenarien in den fehlerfreien Daten enthalten und somit nicht zu melden sind, und welche Verläufe der relativen Fehler bei fehlerhaften Daten vorliegen, die deutlich unterschiedlich zu den fehlerfreien Szenarien sind und somit auf einen Fehler hinweisen.

Die Parameter der einzelnen Kriterien sollten so bestimmt werden, dass für die fehlerfreien Trainingsdaten kein Messpunkt als fehlerhaft gemeldet wird. Trotzdem sollte das Diagnoseverfahren sensibel gegenüber Fehlern in den Messpunkten sein. Es ist daher die richtige Abstimmung der Grenzen zu finden. Das bedeutet, dass die Grenzen nicht zu weit eingestellt werden dürfen, da sonst zu viele Fehler nicht als solche erkannt werden (*false negative*). Andererseits muss man darauf achten die Grenzen nicht zu eng zu wählen, was zur fälschlichen Fehlermeldung von fehlerfreien Punkten führen würde (*false positive*).

Die Parameter werden anhand sogenannter Trainingsdaten eingestellt, die äquivalent zu den Daten der Evaluierung sind. Diese Parameter betreffen etwa die Wahl der Faktoren für die unterschiedlich großen Vertrauensbereiche, die Gruppengrößen beim 2. und 3. Kriterium und die maximale Anzahl an Werten in Folge außerhalb der Grenze bzw. an Gruppenmittelwerten signifikant unterschiedlich zur Referenz (0 oder Gesamtmittelwert). Die folgende Tabelle gibt eine Übersicht über die Kriterien, deren Parameter und wie diese im Rahmen dieser Arbeit eingestellt werden:

Krit.	Betrachtete Größe	Dazugehöriges Konfidenzintervall	Parameterwahl
1	Relative Fehler $\frac{y_i - o_i}{o_i}$	$0 \pm \text{factor } \hat{\sigma} (I_n + x_t(X'X)^{-1}x_t')$ $\hat{\sigma} = RRMSE$	$\text{factor} = [20, 5, 3]$ $\# = [1, 6, 12]$
2	Mean Normalized Error $\frac{1}{n} \sum_{i=1}^n \frac{ f_i - o_i }{o_i}$	$MNE_{Gesamt} + \text{factor } \hat{\sigma}_{are}$	$\text{factor} = 5, \# = 4$ $MP_{Gruppe} = 15$ $MP_{Gesamt} = 90$
3	Mean Normalized Bias $\frac{1}{n} \sum_{i=1}^n \frac{f_i - o_i}{o_i}$	$0 \pm \text{factor } \hat{\sigma}_{are}$	$\text{factor} = 5, \# = 10$ $MP_{Gruppe} = 20$

Tabelle 5.2: Übersicht über die Kriterien und deren Parameter

# bezeichnet dabei in jedem Kriterium die kleinste Anzahl an Werten bzw. Gruppen, die außerhalb des Konfidenzintervalls liegen müssen, um auf einen Fehler in den Messdaten hinzuweisen. Die Parameter mussten dabei teilweise so hoch gewählt werden, um eindeutig sagen zu können, dass die schlechten Prognosen eine Folge fehlerhafter Daten und nicht eines nicht allzu guten Prognoseverfahrens sind. Weiters ist hinsichtlich des 2. und 3. Kriteriums noch ersichtlich, wie viele Messpunkte in einer Gruppe zusammengefasst werden.

## 5.5 Vergleich der Kriterien

Ein Messpunkt wird als fehlerhaft eingestuft, falls mindestens eines der 3 Kriterien dies meldet. Dabei haben alle 3 Kriterien ihre Berechtigung. Große Fehler können und sollen unverzüglich erkannt werden, sowohl im Hinblick auf die Modellqualität als auch auf die praktische Anwendung des Diagnoseverfahrens. Kleine Fehler können zwar auch mit diesen Kriterien entdeckt werden, dabei ist aber die Betrachtung vieler aufeinanderfolgender Messpunkte notwendig. Dies ist jedoch verständlich, da bei sehr kleinen Fehlern die Unterscheidung zwischen tatsächlichem Fehler im System und zufälligem Fehler kaum getätigt werden kann.

Die folgenden Grafiken beschreiben das Zusammenspiel der Kriterien und die Effizienz der einzelnen Methoden. Das Modellgütemaße  $R_{adj}^2$  ist zusätzlich zu den Abbildungen der drei Kriterien ersichtlich. Dies dient zwar nicht zur Messdatendiagnose, soll aber die Auswirkungen von Fehler auf diese Kenngröße verdeutlichen. Weiters wird damit auch ersichtlich, dass dieses Maß durch fehlerhafte Messwerte so weit sinken kann, dass die zuvor erwähnte Schranke von 0.4 unterschritten wird. Zu den davon betroffenen Messpunkten gibt es in Folge keine Vorhersage und somit können diese auch nicht mittels des Diagnoseverfahrens klassifiziert werden.

Folgendes Beispiel zeigt etwa die Notwendigkeit des 1. Kriteriums bzw. dessen großen Vertrauensbereiches. Durch den Sprung in den Messdaten sinkt die Modellqualität rapide. Bereits nach 6 fehlerhaften Messpunkten ist diese so weit gefallen, dass die Schranke von 0.4 unterschritten wird. Würde der große Vertrauensbereich nicht im 1. Kriterium definiert werden (Grau markiert), so würde der Fehler wie in der Abbildung ersichtlich nicht erkannt werden. Dies würde wiederum dazu führen, dass es keine Vorhersagen zu den zeitlich späteren Messpunkten gibt und diese somit nicht durch das Diagnoseverfahren klassifiziert werden können. Sämtliche Kriterien, die für die Erkennung fehlerhafter Messpunkte auf mehr als 6 Messpunkte aufbauen, sprich jedes bis auf das 1. Kriterium mit dem groß definierten Vertrauensbereich, würden gar nicht zum Zuge kommen.

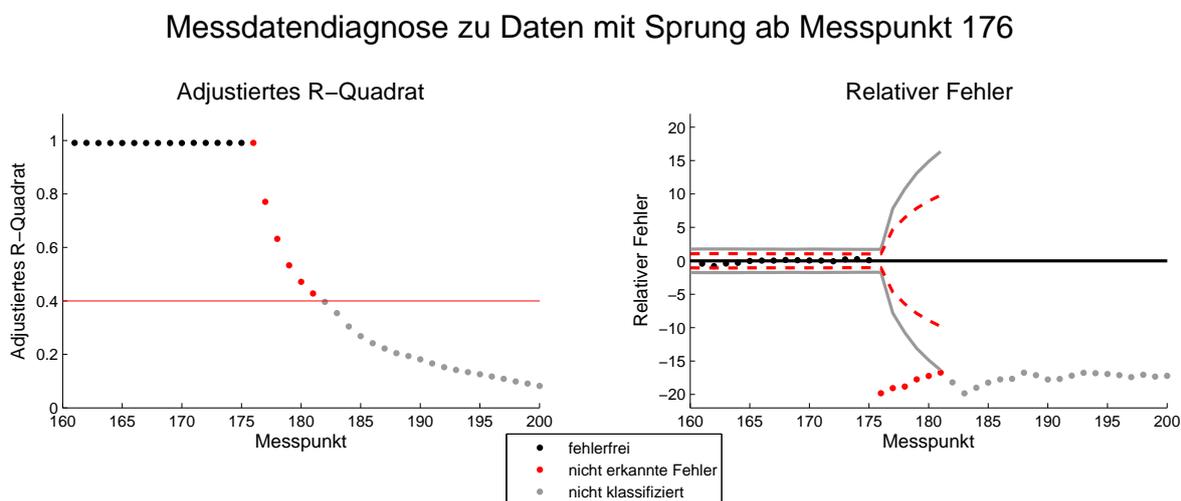


Abbildung 5.6: Notwendigkeit von Kriterium 1 (Großer Vertrauensbereich)

Die Grafik zeigt auch, dass ein Absinken der Modellqualität stets mit einem Wachsen des Vertrauensbereiches einhergeht. Werden Fehler nicht frühzeitig erkannt, so kann dies auch dazu führen, dass der Vertrauensbereich so groß wird, dass die Fehlerdetektion auch an einem späteren Zeitpunkt nicht möglich ist.

Kriterium 2 und 3 zielen, wie bereits erwähnt, eher auf die Erkennung von Mustern ab. Das 2. Kriterium bietet dem 3. Kriterium gegenüber denselben Vorteil wie auch der große Vertrauensbereich im Vergleich zum kleineren: Die fehlerhaften Messpunkte werden früher erkannt und somit nicht in die Modellbildung miteinbezogen.

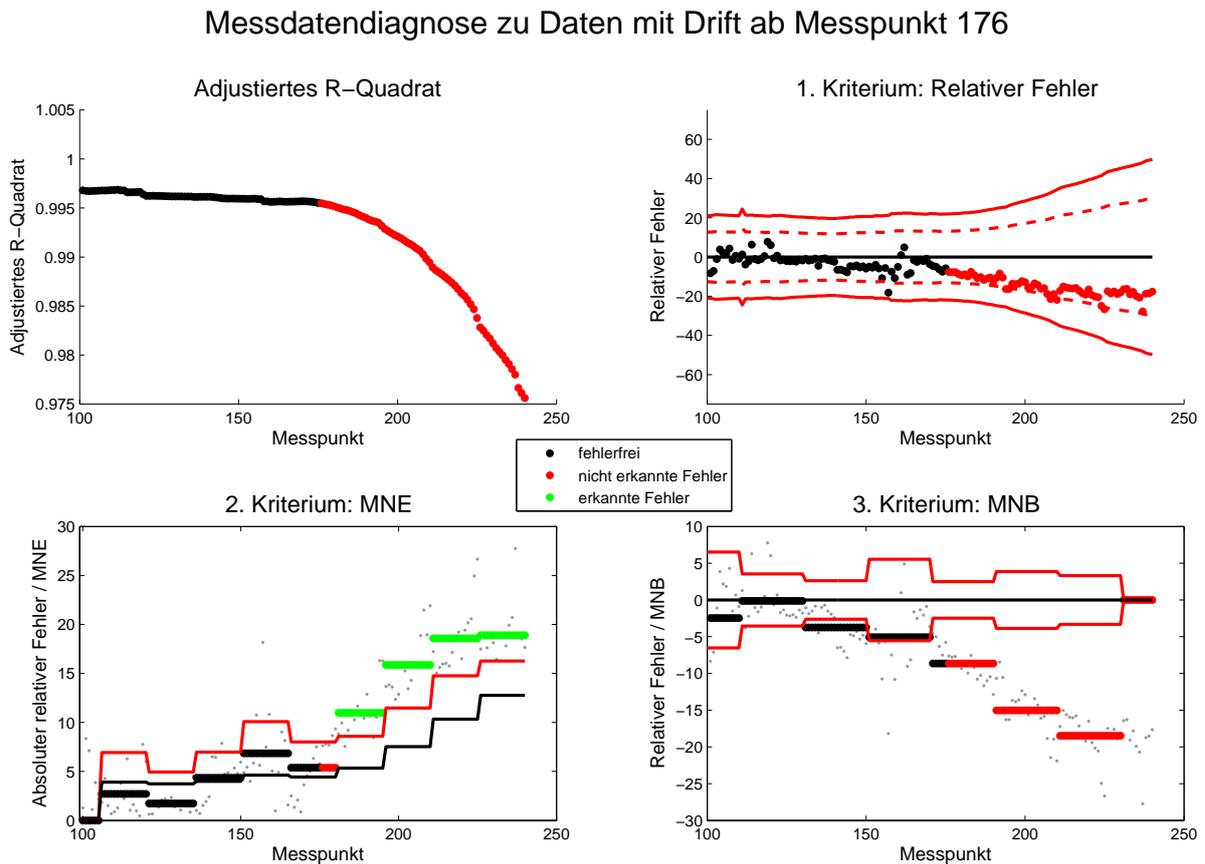


Abbildung 5.7: Vorteil von Kriterium 2

Während der Fehler in den Daten mit dem 2. Kriterium erkannt wird, sind für eine Erkennung durch das 3. Kriterium zu wenig Punkte vorhanden, obwohl eine Unverzerrtheit in den Daten deutlich erkennbar ist. Auch Kriterium 1 kann diesen Fehler in den Daten nicht aufdecken, da dieser eine zu geringe Auswirkung auf den relativen Fehler hat und dieser somit in allen definierten Vertrauensbereichen liegt.

Das sensibelste Kriterium ist Kriterium 3. Dieses kann oft in Situationen Fehler detektieren, an denen die Kriterien 1 und 2 scheitern. Ein Beispiel dafür sind schwache Drifts, wie in der folgenden Abbildung dargestellt.

Messdatendiagnose zu Daten mit Drift ab Messpunkt 176

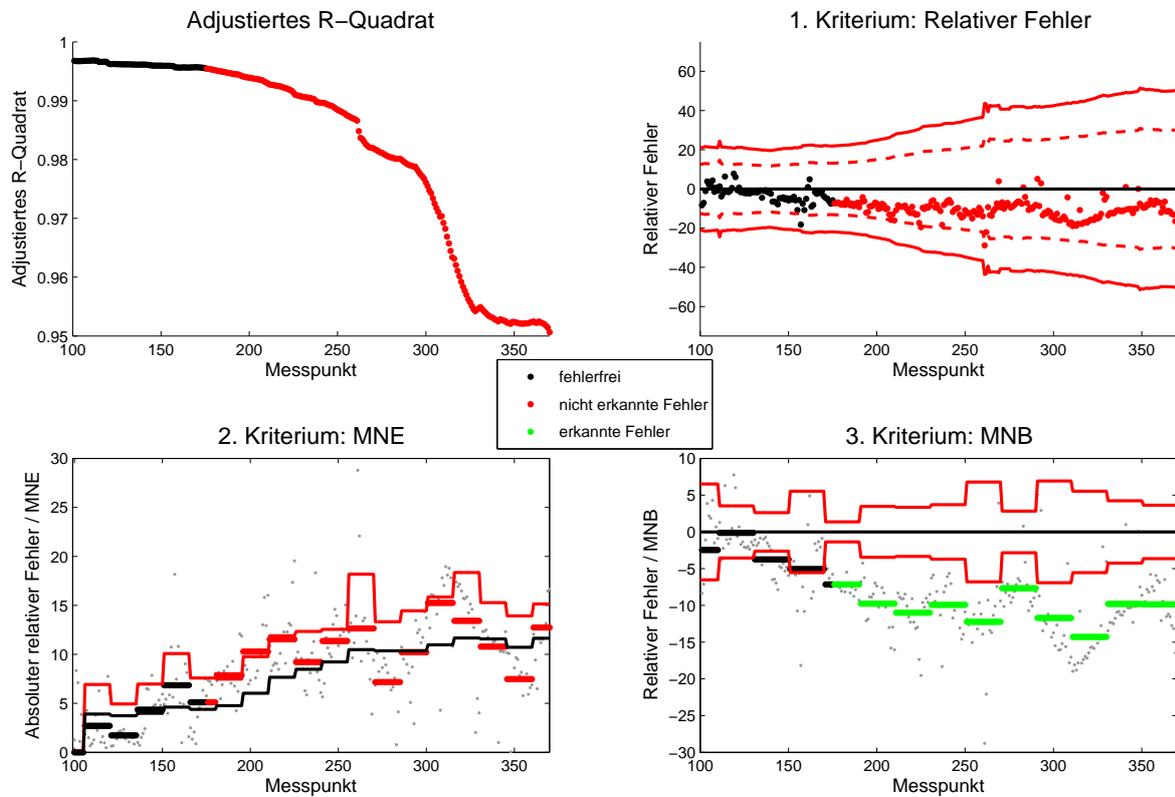


Abbildung 5.8: Vorteil von Kriterium 3

Die Messpunkte scheinen anhand der Betrachtung des Vertrauensbereiches und der absoluten Gruppenmittelwerte nicht auffällig zu sein. Anhand der Kennzahl für die Unverzerrtheit der Daten, dem gruppenweise berechneten MNB, kann man deutlich sehen, dass es über viele Messpunkte hinweg zu einer signifikanten Unterschätzung durch die Prognosen kommt, weswegen der relative Fehler der Prognose weit unter Null liegt.

# 6 Evaluierung

Um eine Aussage über die Güte der Messdatendiagnose zu geben, muss diese in einem geeigneten System implementiert und weiters über Testversuche evaluiert werden. Um die Ergebnisse der Evaluierung quantitativ über Kennzahlen auszudrücken werden bereits bekannte Maßzahlen für solche Klassifizierverfahren angewandt. Da es bereits einen ersten Entwurf von Modellen und Diagnoseverfahren an der LEC GmbH gibt, ist es auch von Interesse, ob diese neu entwickelte Methode eine Verbesserung darstellt. Auch dieser Vergleich kann mittels der Kennzahlen geschehen. Bei der Evaluierung wird dabei sowohl nach Fehlerisolation und Fehlererkennung unterschieden als auch nach Fehlern in den unterschiedlichen Zielgrößen. Ziel ist es dabei zu sehen, bei welchen fehlerhaften Responsevariablen vorhandene Fehler eher entdeckt werden und bei welchen es auch gelingt diesen zu isolieren und somit die fehlerhafte Messgröße selbst zu identifizieren.

## 6.1 Alternative Methode

Die in dieser Arbeit entwickelten Modelle und Messdatendiagnosen können mit einer bereits vorhandenen Methode verglichen werden. Diese beruht zwar auf demselben Prinzip der Prognosewerte, unterscheidet sich aber sowohl in den Modellen als auch im Diagnoseverfahren. Die Basis der vorhandenen Modelle bildet etwa eine Hauptkomponentenanalyse. Diese Analyse sowie die Modellbildung werden in jedem Schritt erneut durchgeführt. Das entsprechende Messdatenklassifizierverfahren ist etwas einfacher. Es basiert lediglich nur darauf, dass eine gewisse Anzahl aufeinander folgender Punkte außerhalb eines Vertrauensbereiches als fehlerhaft gemeldet wird. Das hier entwickelte Diagnoseverfahren stellt somit eine Erweiterung dieses Verfahrens dar.

Die zentrale Idee der Hauptkomponentenzerlegung, im Englischen Principle Component Analysis (PCA) genannt, ist es eine große Anzahl von Messvariablen durch eine geringere Zahl von künstlichen Variablen, die sogenannten „Hauptkomponenten“, anzunähern und dabei einen möglichst großen Anteil der Variation in den Variablen zu erhalten. Es wird dabei durch Transformierung der Daten über Linearkombinationen

eine neue Menge an Variablen erzeugt, die den großen Vorteil bieten, dass sie unkorreliert zueinander sind. So wird dies etwa in JOLLIFFE (1986) beschrieben, wo eine ausführliche Beschreibung dieses Themas gefunden werden kann.

In der Regressionsanalyse findet diese Zerlegung vor allem dann Anwendung, wenn man das Problem der sogenannten Multikollinearität lösen möchte. Diese ist etwa dann vorhanden, wenn Teilmengen der Prädiktoren stark miteinander zusammenhängen. Die Hauptkomponentenregression schafft dies in dem die Prädiktoren selbst durch deren Hauptkomponenten ersetzt werden, die wie bereits erwähnt unkorreliert sind. Werden alle Hauptkomponenten in das Modell inkludiert, so erhält man dasselbe Ergebnis wie mit Hilfe der Kleinst-Quadrate-Schätzung. Daher bedarf es der Wahl einer Teilmenge, zu der es unterschiedliche Strategien gibt. Für nähere Details dazu möchte der Leser in JOLLIFFE (1986) nachschlagen.

Bezüglich der praktischen Umsetzung sei angemerkt, dass vor der Zerlegung eine Standardisierung der Daten stattfindet, da die Maßeinheiten variieren und sich die Varianzen der einzelnen Messgrößen stark unterscheiden. Wie bereits erwähnt wird diese Art der Modellbildung, im Gegensatz zu den in allen Messpunkten spezifizierten Modellen dieser Arbeit, in jedem neuen Messpunkt erneut durchgeführt. Außerdem wird dies auf Basis aller vergangenen Messpunkte getan, sofern diese sich etwa durch unterschiedliche Bauteile des Motors nicht deutlich unterscheiden. Die Vorhersage selbst wird wiederum nur dann berechnet, wenn die Modelle im Modellgütemaß eine gewisse Schranke überschreiten. Die für den Vergleich verwendete Methode benutzt als Schranke den sehr hohen Wert von 0.95.

Das Diagnoseverfahren ist, wie erwähnt, eine Vereinfachung der in dieser Arbeit vorgestellten Methode. Es bezieht sich auch auf den relativen Fehler der Vorhersage, der tatsächliche Wert wird aber nur in Hinsicht auf ein Kriterium als fehlerfrei oder fehlerhaft klassifiziert. Dieses entspricht dem 1. Kriterium dieser Arbeit, wobei jedoch nur ein Vertrauensbereich definiert wird. Im Zuge der Parameterwahl wird der Faktor des Vertrauensbereiches auf 4 eingestellt und die Anzahl der relativen Fehler in Folge außerhalb der Grenze darf 5 nicht überschreiten. Wiederum werden bereits als fehlerhaft gemeldete Punkte nicht in die Modellbildung miteinbezogen.

## 6.2 Kennzahlen zur Bewertung

Um die Güte der gewählten Modelle kombiniert mit der fortgeschrittenen Messdaten-diagnose bewerten zu können, finden gewisse Kennzahlen Gebrauch. Diese können auch für den Vergleich mit dem alternativen Ansatz angewandt werden und sollen numerisch beschreiben, wie sensibel die Methoden gegenüber Fehlern sind und ob es zu Falschdetektionen, bei denen ein in Wirklichkeit fehlerfreier Messpunkt als fehlerfrei gemeldet wird, kommt. Sofern eine Prognose für den Messpunkt getätigt wird und die Wahrheit über die Güte des Messpunktes bekannt ist (fehlerhaft oder nicht fehlerhaft), kann dieser in eine der 4 Kategorien eingeteilt werden:

Wahrheit	Entscheidung	
	Fehler	kein Fehler
Fehler	true positive	false negative
kein Fehler	false positive	true negative

Tabelle 6.1: Kategorien eines klassifizierten Messpunktes

Ein *true positive* Punkt ist ein tatsächlicher Fehler, der auch richtig detektiert wird. Wird ein Messpunkt als fehlerhaft detektiert, obwohl dieser in Realität fehlerfrei ist, so spricht man von einem *false positive* Punkt. Dazu gibt es im Gegenteil *false negative* Punkte, die zwar fehlerhaft sind, dies aber nicht diagnostiziert wird. Wie in WOLFF (2010) kann man auch von einem korrekten Alarm, einem fehlenden Alarm und einem Fehlalarm sprechen. Summiert man über alle Messpunkte einer Messreihe bzw. eines Datensatzes auf, so können zur Evaluierung folgende bekannte Maßzahlen (siehe RUNKLER (2015)) verwendet werden:

$$\text{Genauigkeit (precision)} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

$$\text{Trefferquote (recall)} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$$

Ein Klassifikator ist umso besser, je näher diese Kennzahlen bei 1 sind. Man kann anhand der Formeln erkennen, dass desto weniger *false positive* Punkte es gibt, desto höher ist die Genauigkeit. Weiters ist die Trefferquote höher, je weniger *false negative* Punkte detektiert werden. Werden diese beiden Maße zur Bewertung von Klassifikatoren herangezogen, so kommt es auf die Anwendung an, ob eher eine hohe Genauigkeit oder eine hohe Trefferquote gewünscht ist. Ist es entscheidend, die Anzahl der Fehlalarme zu minimieren, so ist das Ziel, einen hohen Genauigkeitswert zu erhalten. Will man jedoch möglichst keine fehlenden Detektionen bzw. Alarme haben, so sollte auf

eine möglichst hohe Trefferquote abgezielt werden. In diesem Fall ist es wünschenswert, auch im Interesse des Anwenders, eine möglichst hohe Genauigkeit zu erreichen. Daher sollten die Grenzen nicht allzu eng eingestellt werden, um fälschlicherweise fehlerhaft gemeldete Punkte zu vermeiden.

Da in diesem Fall mehrere Messgrößen gleichzeitig nach Fehlern untersucht werden und es für den Anwender wünschenswert ist zu wissen, welche konkrete Messgröße fehlerhaft ist, kann man hier weiters noch unterscheiden zwischen Fehlererkennung und Fehlerisolation (siehe dazu etwa WOLFF (2010)). Bei der Fehlererkennung bzw. Fehlerdetektion geht es darum, zu erfahren, ob in irgendeiner der beobachteten Messgrößen ein Fehler aufgetreten ist. Somit wird eine Antwort auf die Frage „Ist das System fehlerhaft?“ gegeben. Hingegen geht es bei der Fehlerisolation darum, zu erfahren, ob nur in der tatsächlich fehlerhaften Messgröße ein Fehler detektiert wurde. Man will hier also genau wissen, welchen Teil des Systems der Fehler betrifft.

Bei der Einteilung der Messpunkte nach den 4 Kategorien werden bei der Fehlererkennung die Messpunkte aller Messgrößen vereint betrachtet. Ein Messpunkt wird dann in der Entscheidung als fehlerhaft eingestuft, wenn für mindestens eine Messgröße an diesem Messpunkt ein Fehler detektiert wurde. Bei der Fehlerisolation werden die Messpunkte aller Messgrößen einzeln betrachtet und klassifiziert. Bei der anschließenden Analyse werden die zuvor genannten Kennzahlen sowohl für die Fehlerdetektion als auch für die Fehlerisolation berechnet und ausgegeben, damit auch beurteilt werden kann, in wie weit die tatsächlich fehlerhafte Messgröße identifiziert werden kann. Je größer die Genauigkeit und Trefferquote der Fehlerisolation, desto weniger Einfluss haben Fehler einer Messgröße auf die anderen Messgrößen und desto einfacher kann die tatsächlich fehlerhafte Messgröße bestimmt werden.

Weiters ist in der Praxis noch das richtige Markieren der fehlerhaften Punkte, also das richtige Klassifizieren von Bedeutung. Werden etwa Muster von aufeinanderfolgenden relativen Fehler betrachtet, so kann man erst beim letzten Punkt bzw. bei der letzten Gruppe sicher sein, dass es sich hier um einen Fehler handelt, da kleinere Gruppen anhand der Kriterien als fehlerfrei gesehen werden. Bei der vorgenommenen Evaluierung wurde dies so umgesetzt, dass nur der letzte Punkt beim 1. Kriterium oder die letzte Gruppe eines Musters beim 2. und 3. Kriterium als fehlerhaft gekennzeichnet wurden.

## 6.3 Prinzip der Evaluierung

Um die Methoden mit den genannten Kennzahlen evaluieren und die Messpunkte, wie im vorhergehenden Abschnitt beschrieben, klassifizieren zu können, bedarf es Messreihen, deren Messpunkte gezielt manipuliert wurden, sodass die Wahrheit „Fehler oder kein Fehler“ bereits vorab bekannt ist. Für die Evaluierung sind also Datensätze notwendig, bei denen künstlich Fehler eingebaut worden sind, sogenannte „Testcases“.

Als Grundlage für diese „Testcases“ werden 3 Datensätze mit 526, 1013 und 1016 Messpunkten, die innerhalb eines Datensatzes alle von derselben Konfiguration stammen, verwendet. Die unterschiedliche Länge der Datensätze ist dabei wichtig, da somit die Schwäche der Kriterien, die auf vielen Messpunkten aufbauen, ersichtlich wird.

Im Anschluss werden damit Datensätze generiert, bei denen in jeweils einer Messgröße (eine der neun Zielgrößen) künstlich ein Fehler eingebaut wird. Dabei werden sowohl Drifts, sprich kontinuierliche Steigerungen des Betrages des Fehlers ab einem gewissen Messpunkt bis zum Ende des Datensatzes, als auch anhaltende Sprünge, also ein Änderung gewisser Höhe ab einem bestimmten Messpunkt bis zum Ende gleichbleibend, auf die Originaldaten aufgetragen. Die Änderung ist dabei in Relation zum gemessenen Originalwert der geänderten Größe zu sehen. Genauer wird diese Fehleraufprägung im Folgenden erläutert.

Sei dazu  $i$  der erste künstlich fehlerhafte Messpunkt und  $l$  der letzte Messpunkt des Datensatzes. Ein anhaltender, positiver Sprung von z.B. 50% im Verhältnis zur Originalmessreihe kann mathematisch folgendermaßen formuliert werden:

$$y_{fehlerhaft}[i, l] = y_{original}[i, l] \times (1 + 0.5).$$

Die sieht für ein Beispiel von 10 Messpunkten und einem ab dem 6. Messpunkt aufgetragenen Sprung von 50% derart aus:

Messpunkt	1	2	3	4	5	6	7	8	9	10
Änderung	0%	0%	0%	0%	0%	50%	50%	50%	50%	50%
Multiplikator	1.00	1.00	1.00	1.00	1.00	1.50	1.50	1.50	1.50	1.50

Tabelle 6.2: Beispielhafte Fehleraufprägung für einen Sprung

Auch die Fehleraufprägung bei einem schleichenden Fehler bzw. Drift von z.B. 3% über 10 Messpunkte im Verhältnis zur Originalmessreihe kann mathematisch formuliert werden. Sei dazu wiederum  $i$  der erste künstlich fehlerhafte und  $l$  der letzte Messpunkt des Datensatzes. Sei weiters  $1 : (l - i)$  die Abkürzung für den Vektor  $[1, 2, \dots, l - i -$

## 6 Evaluierung

$1, l - i]$ . Ein Drift kann dann derart dargestellt werden:

$$y_{fehlerhaft}[i, l] = y_{original}[i, l] \times (1 + (1 : (l - i)) \times 0.003).$$

Wiederum wird dies mit einem Beispiel von 10 Messpunkten erklärt, bei denen ab dem 6. Messpunkt ein Drift von 0.3% über 2 bzw. 3% nach 10 Messpunkten eingebaut wird:

Messpunkt	1	2	3	4	5	6	7	8	9	10
Änderung	0%	0%	0%	0%	0%	0.3%	0.6%	0.9%	1.2%	1.5%
Multiplikator	1.00	1.00	1.00	1.00	1.00	1.003	1.006	1.009	1.012	1.015

Tabelle 6.3: Beispielhafte Fehleraufprägung für einen schleichenden Fehler

Anhand der Grafik wird der Unterschied zwischen den beiden Fehlerarten deutlich:

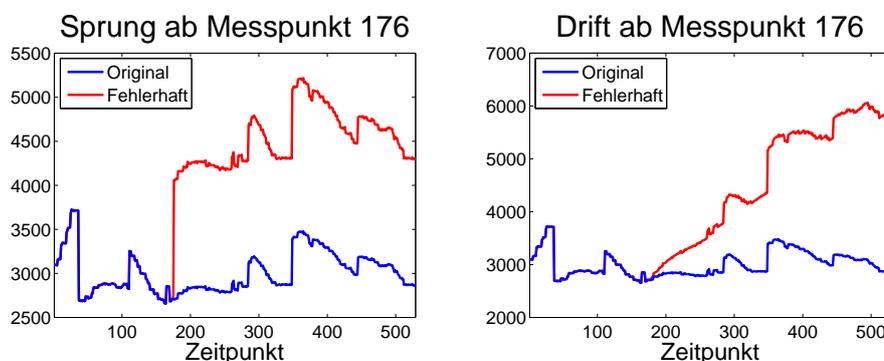


Abbildung 6.1: Beispielhafte Aufprägungen der beiden Fehlerarten

Es kann weiters unterschieden werden nach der Stärke der Aufprägung, also der unterschiedlichen Höhe der aufgetragenen Abweichung zur Originalmessreihe (Angabe in Prozent), und nach der Stelle des Beginns des Fehlers im Datensatz (jeweils 3 verschiedenen Stellen). Jeder Testcase wird durch die Kennzahlen Genauigkeit und Trefferquote bzgl. Fehlerisolation und Fehlerdetektion evaluiert. Über die unterschiedlichen Datensätze und Stellen der Auftragungen wird gemittelt. Ziel ist es zu erkennen, bei welchen Messgrößen der Anwender mehr bzw. weniger Acht geben muss, da die Fehlerdiagnose weniger bzw. mehr gute Ergebnisse liefert.

## 6.4 Ergebnisse

### 6.4.1 Performance der entwickelten Methode

Es werden nun die Resultate der Evaluierung für die in dieser Arbeit entwickelte Methode gezeigt. Es wird dabei nach den unterschiedlichen Messgrößen unterschieden, in

die ein Fehler eingebaut wurde. Die Ergebnisse werden grafisch präsentiert. Die beiden Fehlerarten, Sprung und Drift, werden dabei getrennt betrachtet. Außerdem wird in den Plots selbst nach Drift- bzw. Sprunghöhe und Fehlererkennung bzw. Fehlerisolation unterschieden. Die Ergebnisse der Evaluierung der entwickelten Methode bezüglich eines Sprunges sind in der folgenden Grafik dargestellt:

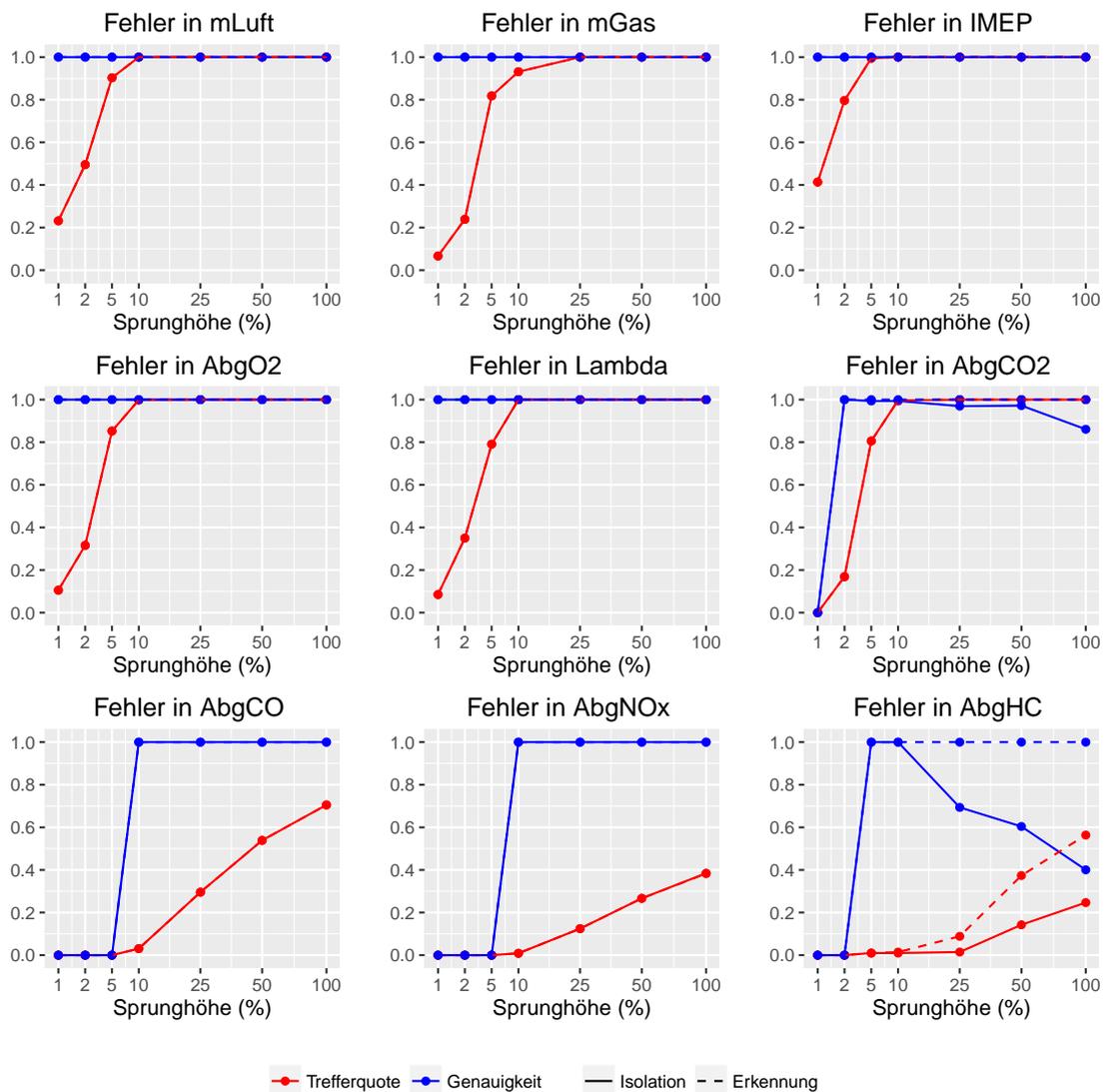


Abbildung 6.2: Auswertung der Kennzahlen zum Sprung je Messgröße

Sobald die Trefferquote über Null liegt, ist die Genauigkeit der Fehlerisolation bei nahezu allen Messgrößen bei 100%. Dies liegt wohl daran, dass sowohl die Parameter der Diagnose gut eingestellt sind als auch die fehlerhaften Punkte richtig markiert werden. Ausnahmen sind die Messgrößen AbgCO2 und AbgHC. Da diese als Prädiktoren in anderen Modellen verwendet werden, führt ein Fehler in ihnen zu fehlerhaften Vorhersagen

## 6 Evaluierung

in den anderen Modellen und somit zu *false positive* Messpunkten im Hinblick auf die Fehlerisolation. Daraus folgt auch, dass die Genauigkeit der Fehlererkennung jedoch auch bei diesen beiden Messgrößen bei 100% liegt. Dies hat jedoch auch einen Vorteil. So werden etwa im Zuge der Fehlererkennung für die Größe AbgHC mehr fehlerhafte Punkte als solches detektiert, wodurch die Trefferquote für die Fehlererkennung höher ist als für die Fehlerisolation. Die Methode zeigt sehr gute Ergebnisse selbst bei sehr kleinen Veränderungen. Ausgenommen sind dabei die Messgrößen AbgCO, AbgNOx und AbgHC. Hier erreicht die Trefferquote auch bei sehr großen Sprunghöhen nicht die 100%. Bevor auf diese Tatsache näher eingegangen wird, wird zuerst noch ein Blick auf die Ergebnisse geworfen, die für den künstlich eingebauten schleichenden Fehler resultieren.

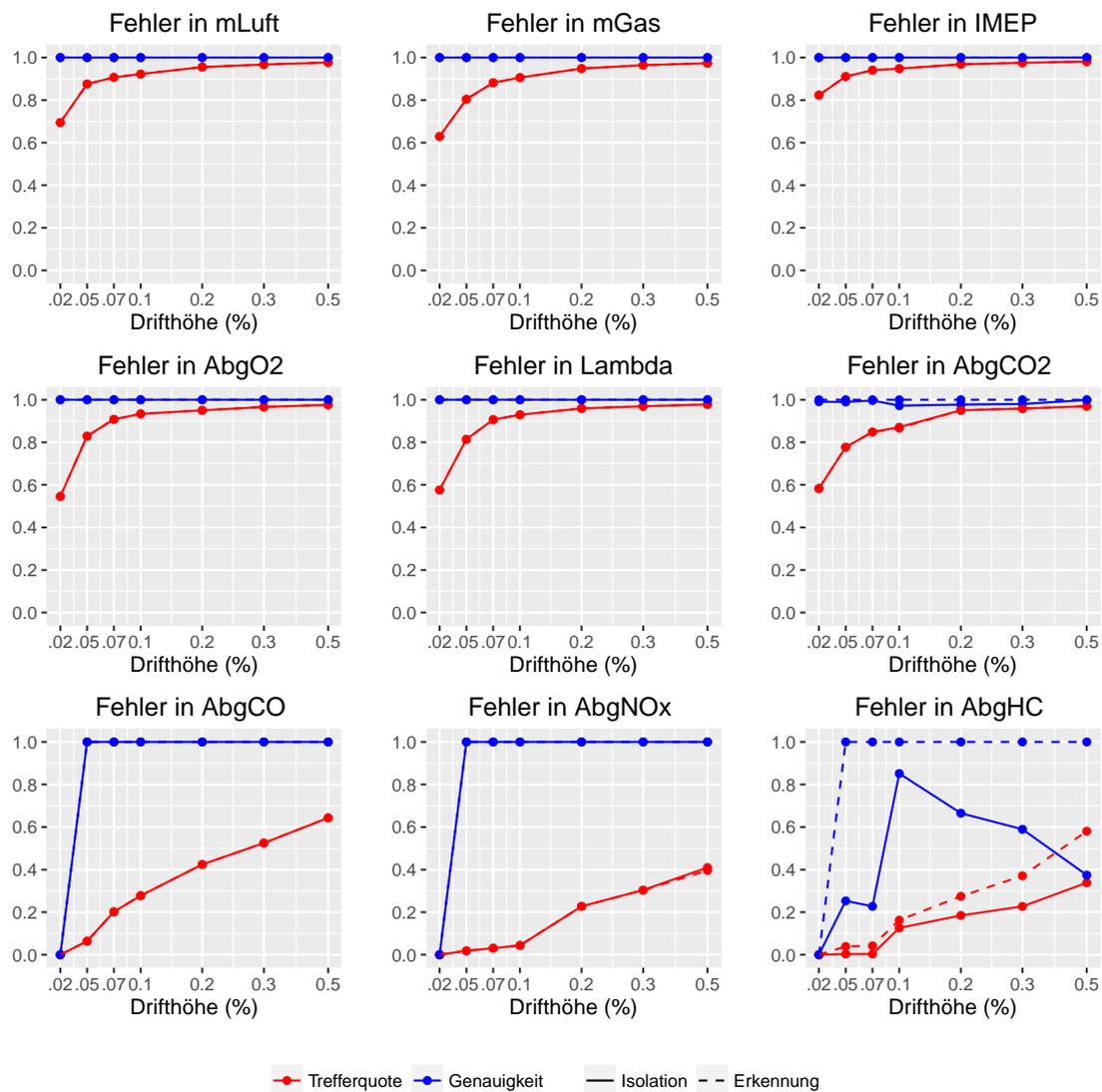


Abbildung 6.3: Auswertung der Kennzahlen zum Drift je Messgröße

Auch hier gibt es bis auf die beiden Ausnahmen  $\text{AbgCO}_2$  und  $\text{AbgHC}$  eine Genauigkeit von 100% in der Fehlerisolierung. Insgesamt zeigt sich ein ähnliches Bild wie beim Einbau eines künstlichen Sprunges. Sehr hohe Trefferquoten bei den Messgrößen  $\text{mLuft}$ ,  $\text{mGas}$ ,  $\text{IMEP}$ ,  $\text{AbgO}_2$ ,  $\text{Lambda}$  und  $\text{AbgCO}_2$  und eher niedrigere Werte bei den restlichen Messgrößen. Um dieses Phänomen und auch das der kleineren Genauigkeit bei als Prädiktoren verwendeten Zielgrößen näher zu analysieren, folgen nun Scatterplots, bei denen diese Kennzahlen für einen sehr hohen Sprung bzw. Drift in Verbindung mit den Variationskoeffizienten von Abschnitt 3.1 gebracht werden.

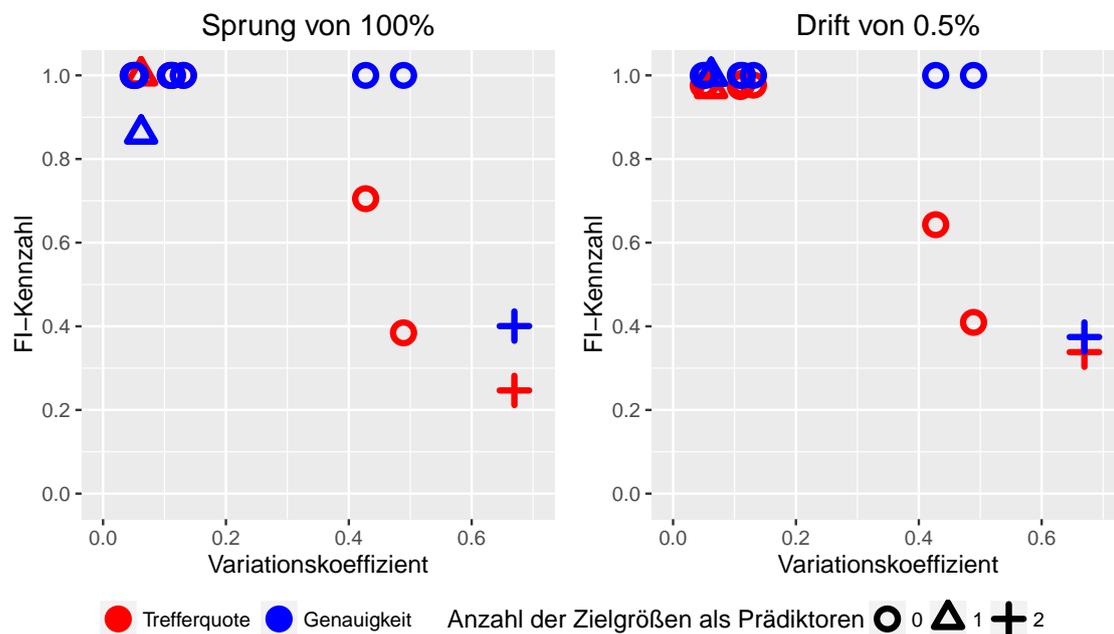


Abbildung 6.4: Scatterplot: Kennzahlen bzgl. der Fehlerisolierung gegen Variationskoeffizienten der Zielgrößen

Bei der Betrachtung dieser Grafik wird die Ursache der geringen Trefferquote bei Fehlern in den Messgrößen  $\text{AbgCO}$ ,  $\text{AbgNO}_x$  und  $\text{AbgHC}$  deutlich. Es ist ganz klar ersichtlich, dass der hohe Variationskoeffizient dieser Messgrößen eine wichtige Rolle spielt. Das bedeutet, dass die Variation dieser Messgrößen zu groß ist, um fehlerhafte Variationen zu erkennen. Ein vorliegender Fehler geht somit im sogenannten ‘‘Weißen Rauschen’’, sprich dem zufälligen Fehler der Messgröße, unter und kann deswegen nicht von der Diagnose als fehlerhaft erkannt werden. Mit Hilfe der Symbole kann man weiters noch erkennen, dass die Genauigkeit bei jenen Messgrößen nicht bei 100% liegt, die zur Modellbildung bzw. Vorhersage verwendet werden, wie bereits zuvor erläutert.

Insgesamt kann für den Anwender gefolgert werden, dass die Diagnose für die Messgrößen  $\text{mLuft}$ ,  $\text{mGas}$ ,  $\text{IMEP}$ ,  $\text{AbgO}_2$ ,  $\text{Lambda}$  und  $\text{AbgCO}_2$  sehr gute Ergebnisse liefert und

somit in der Anwendung selbst äußerst zuverlässig scheint, da sogar kleine Änderungen meist erkannt werden. Bei den Messgrößen  $\text{AbgCO}$ ,  $\text{AbgNO}_x$  und  $\text{AbgHC}$  hingegen werden systematische Fehler oft durch den zufälligen Fehler der Messgröße selbst überlagert und somit kaum detektiert. Es müssen sich somit der tatsächliche Messwert und der Prognosewert stark voneinander unterscheiden, um eine signifikante Abweichung vom Systemverhalten zu erkennen. Der Wunsch des Anwenders keine Falschdetektionen zu haben ist zumindest bezüglich der Fehlererkennung erfüllt.

### 6.4.2 Vergleich der beiden Methoden

Es folgen nun die Ergebnisse zum Vergleich der beiden Methoden. Die in dieser Arbeit entwickelte Methode wird dabei mit Methode „neu“ und die bereits vorhandene Methode mit Methode „alt“ bezeichnet. Dabei wird für jede Methode über die 9 Messgrößen, in die künstlich ein Fehler eingebaut wurde, gemittelt. Weiterhin ist aber die Unterscheidung nach der Stärke der Änderung und nach der Art des Fehlers von Interesse.

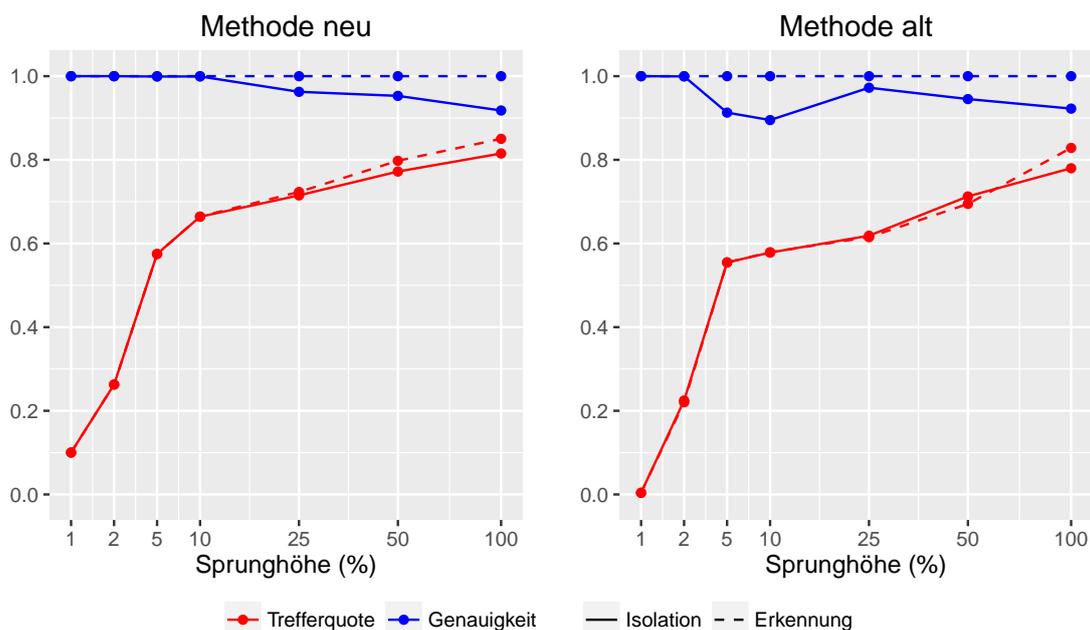


Abbildung 6.5: Auswertung der Kennzahlen zum Sprung der beiden Varianten

Hinsichtlich der Fehlererkennung ist die Genauigkeit der beiden Methoden gleich hoch, diese liegt nämlich bei 100%. Das bedeutet, dass auch die alternative Methode nur fehlerhafte Punkte in einer Messgröße detektiert, wenn in einer anderen oder derselben tatsächlich ein Fehler vorliegt. Bezüglich der Trefferquote stellt die neu entwickelte Methode bei allen Sprunghöhen eine leichte Verbesserung dar.

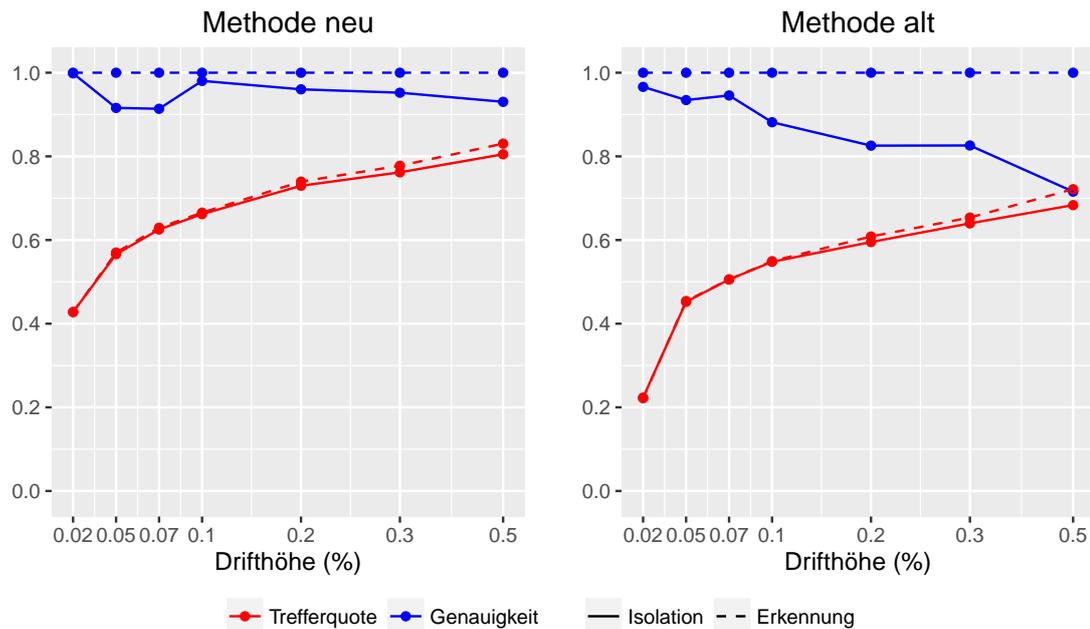


Abbildung 6.6: Auswertung der Kennzahlen zum Drift der beiden Varianten

Auch beim Einbau von künstlichen Drifts zeigt sich das gute Verhalten beider Varianten in Hinsicht auf die Trefferquote der Fehlererkennung. Bei der Fehlerisolation stellt die neue Methode, zumindest bei größeren Änderungen, eine Verbesserung in der Genauigkeit dar. Besonders im Hinblick auf die Trefferquote kann eine signifikante Steigerung mit Hilfe der neuen Methode erkannt werden. Für nicht allzustark ausgeprägte, schleichende Fehler dürfte das vor allem an den Kriterien 2 und 3 liegen.

Zum Abschluss muss im Rahmen des Vergleichs der beiden Methoden noch erwähnt werden, dass es bei der alten Methode vor allem bei der Messgröße  $\text{AbgNO}_x$  dazu kommt, dass keine Modelle mit entsprechend hohem Bestimmtheitsmaß gefunden und somit keine Vorhersagen gebildet werden. Dazu könnte man sich etwa die Vorhersagerate ansehen, bei der die Anzahl der tatsächlich vorhergesagten und somit klassifizierten Messpunkte in Relation zur gesamten Anzahl an Messpunkten eines Datensatzes gesetzt wird. Diese liegt bei der neu entwickelten Methode bei etwa 96%, bei der alten Methode hingegen nur bei etwa 78%. Dabei muss weiters erwähnt werden, dass die 100% von beiden Methoden nicht erreicht werden können, da die Messdatendiagnose und somit die Klassifizierung erst dann gestartet wird, wenn ausreichend Punkte zur Modellbildung vorhanden sind.



## 7 Resümee und Ausblick

**Datenlage.** Die Notwendigkeit der Aufteilung der Datenpunkte nach den Motorbauteilen Nockenwelle und Kolben kann durch Residuenplots und Verwendung der Indikatorvariablen vor allem für die Zielgrößen  $m_{Luft}$ ,  $m_{Gas}$  und  $IMEP$  bestätigt werden. Durch die hohe Variationsbreite mancher Zielgrößen gelingt dem Diagnoseverfahren in diesen Fällen nur bei sehr großen Fehlern zumindest teilweise eine Fehlererkennung. Dies trifft etwa bei einem Drittel der Zielgrößen zu.

**Modelle.** Die Modelle, gefunden durch eine Mischung aus automatischem Vorgehen mittels Schrittweiser Regression und manueller Auswahl der Prädiktoren, liefern eine akzeptable Anpassungsgüte. Ein Vorteil bietet sich durch die Modellierung des nicht linearen Zusammenhangs der Messgröße  $AbgNO_x$  zu den anderen über Transformationen der Prädiktoren. Ein Nachteil hingegen ist allerdings, dass einige Modelle andere Zielgrößen als Prädiktoren haben. Dies hat zur Folge, dass im Hinblick auf die Evaluierung der Methode die Genauigkeit der Fehlerisolation sinkt. Bei Messgrößen, die über lange Zeitabschnitte konstant gehalten werden, wird die Modellbildung erschwert, da dort kaum eine zu erklärende Varianz vorhanden ist.

**Diagnoseverfahren.** Alle drei entwickelten Kriterien zeigen ihre Vorteile in gewissen Situationen. Um geringe, fehlerhafte Veränderungen in den Messpunkten zu detektieren, bedarf es der Analyse einer relativ großen Anzahl an Datenpunkten. Die vom Anwender gewünschten 100% Genauigkeit werden zumindest bzgl. der Fehlererkennung erreicht. Außerdem stellt die neu entwickelte Methode besonders bei schleichenden Fehlern eine Verbesserung gegenüber der bereits bekannten Methode im Hinblick auf die Trefferquote dar.

**Mehrere fehlerhafte Größen.** Bisher wurde nur in je einer Zielgröße ein simulierter Fehler eingebaut. Es wäre durchaus interessant zu sehen, wie sich das Diagnoseverfahren verhält, wenn das System in einer erklärenden Größe bzw. in mehreren Größen gleichzeitig fehlerhaft ist, insbesondere in Hinsicht auf die Fehlerisolation. Weiters könnte ein noch größerer Datenumfang für den Anwender aufschlussreich sein, um zum Beispiel Analysen zu Messdaten mit einer Drehzahl-Variation durchzuführen.

**Evaluierung unter realen Umständen.** Von Interesse ist weiters sicherlich eine Evaluierung der Methode unter realen Umständen. Dies könnte sich jedoch als schwierig erweisen, da man dann nicht in jeder Situation weiß, ob in Wahrheit ein fehlerhafter Messpunkt vorliegt. Wird zum Beispiel ein tatsächlich vorliegender Fehler weder vom Diagnoseverfahren noch vom Prüfstandspersonal oder anderen Überprüfungstechniken erkannt, so kann dieser *false negative* Messpunkt nicht als solcher klassifiziert werden, sondern wird fälschlicherweise als *true negative* Punkt gezählt.

**Messgrößen-System.** Im Hinblick auf die Modelle wäre ein sogenanntes Stellgrößen-Zielgrößen-System ideal, um so den Fehler tatsächlich isolieren und die fehlerhafte Messgröße exakt bestimmen zu können. Ist dies nicht möglich, so muss eine andere Technik gefunden werden, um bei einem detektierten Fehler in mehreren Größen entscheiden zu können, in welcher Messgröße der Fehler tatsächlich vorliegt.

**Ausblick.** Insgesamt wäre eine Generalisierung dieser Diagnose auf beliebige Motortypen bzw. Motorprüfstände wünschenswert. Da es in der Literatur schon andere Arten der Fehlererkennung, wie etwa über Grenzwertermittlungen, gibt, wäre auch eine Kombination unterschiedlicher Methoden denkbar, um eine ausreichend gute Lösung für die Problematik der frühzeitigen Erkennung systematischer Fehler zu erhalten.

# Literatur

- ALBERS, S.; KLAPPER, D.; KONRADT, U.; WALTER, A.; WOLF, J. (Hrsg.): *Methodik der empirischen Forschung*, Wiesbaden: Deutscher Universitäts-Verlag, 2006.
- BÖGE, A.; BÖGE, W. (Hrsg.): *Handbuch Maschinenbau, 22. Auflage*, Wiesbaden: Springer Vieweg, 2015.
- CLEFF, T.: *Deskriptive Statistik und Explorative Datenanalyse, 3. Auflage*, Wiesbaden: Springer, 2015.
- ECKLE-KOHLER, J.; KOHLER, M.: *Eine Einführung in die Statistik und ihre Anwendungen*, Berlin: Springer, 2009.
- FAHRMEIR, L.; KNEIB, T.; LANG, S.: *Regression, 2. Auflage*, Berlin: Springer, 2009.
- FLOHR, A.: *Konzept und Umsetzung einer Online-Messdatendiagnose an Motorenprüfständen*, Dissertation, Technische Universität Darmstadt, 2005.
- FRIEDL, H.: *Computerstatistik*, Vorlesungsskript, Technische Universität Graz: Institut für Statistik, 2005.
- HACKL, P.: *Einführung in die Ökonometrie*, München: Pearson Studium, 2005.
- HIEBELER, D. E.: *R and MATLAB*, Boca Raton, FL: CRC Press, 2015.
- JOLLIFFE, I. T.: *Principal Component Analysis*, New York: Springer-Verlag, 1986.
- KLEINBAUM, D. G.; KUPPER, L. L.; MULLER, K. E.; NIZAM, A.: *Applied Regression Analysis and Other Multivariable Methods, Third Edition*, Pacific Grove, CA: Duxbury Press, 1998.
- MARÉCAL, V. et al.: *A regional air quality forecasting system over Europe: the MACC-II daily ensemble production*, in: Geoscientific Model Development 8.9 (2015), S. 2777–2813, URL: <http://www.geosci-model-dev.net/8/2777/2015/>.

## Literatur

- MONTGOMERY, D. C.: *Introduction to Statistical Quality Control, 6th Edition*, Hoboken, NJ: John Wiley & Sons, Inc., 2009.
- RUNKLER, T. A.: *Data Mining, 2. Auflage*, Wiesbaden: Springer Vieweg, 2015.
- SEN, A.; SRIVASTAVA, M.: *Regression Analysis*, New York: Springer, 1990.
- STADLOBER, E.: *Angewandte Statistik, Vorlesungsskript*, Technische Universität Graz: Institut für Statistik, 2011.
- VAN BASSHUYSEN, R.; SCHÄFER, F. (Hrsg.): *Handbuch Verbrennungsmotor, 7. Auflage*, Wiesbaden: Springer Vieweg, 2015.
- WEI, T.; SIMKO, V.: *Package ‘corrplot’: Visualization of a Correlation Matrix*, April 2016 - Version 0.77, URL: <https://cran.r-project.org/web/packages/corrplot/corrplot.pdf>.
- WICKHAM, H.: *ggplot2 Elegant Graphics for Data Analysis*, New York: Springer, 2009.
- WIMMER, A.: *Motorenmesstechnik, Vorlesungsskript*, Technische Universität Graz: Institut für Verbrennungskraftmaschinen und Thermodynamik, 2014.
- WIMMER, A.: *Thermodynamik des Verbrennungsmotors, Vorlesungsskript*, Technische Universität Graz: Institut für Verbrennungskraftmaschinen und Thermodynamik, 2012.
- WOLFF, F.: *Konsistenzbasierte Fehlerdiagnose nichtlinearer Systeme mittels Zustandsmengenbeobachtung*, Dissertation, Karlsruher Institut für Technologie, 2010.