Dipl.-Ing. Martin Piffl, Bakk.techn.

Modellbasierte Produktions- und Lebensdauerorientierte Entwicklung am Beispiel eines Dieselmotors

DISSERTATION

zur Erlangung des akademischen Grades
Doktor der technischen Wissenschaften
eingereicht an der

Technische Universität Graz

Betreuer TUGraz:
Univ.-Prof. Dipl.-Ing. Dr.techn. Ernst Stadlober

Institut für Statistik

Betreuer AVL List GmbH:
Dr. Martin Schüßler

Abteilung DEM/PTE

Graz, im August 2015

Dipl.-Ing. Martin Piffl, Bakk.techn.

Model-Based Production and Lifetime-Oriented Development by Means of a Compression Ignition Engine

DOCTORAL THESIS

to achieve the university degree of
Doktor der technischen Wissenschaften
submitted to

Technische Universität Graz

Supervisor TUGraz:
Univ.-Prof. Dipl.-Ing. Dr.techn. Ernst Stadlober

Institute of Statistics

Supervisor AVL List GmbH:
Dr. Martin Schüßler

Department DEM/PTE

Graz, August 2015

Eidesstattliche Erklärung

AFFIDAVIT

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 Dissertation identisch.

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral dissertation.

Graz, am . . . .
                                                                                                    (Unterschrift)

ZUSAMMENFASSUNG

Der Entwicklungsprozess eines KFZ-Antriebstrangs erfordert den kontinuierlichen Einsatz von immer komplexer werdenden Lösungsansätzen für die wesentlichen Bestandteile, den Motor, das Getriebe, elektrische Antriebe, das Abgasnachbehandlungssystem und elektronische Regelungsstrategien. Die dabei entwickelten Qualitätsstandards müssen dabei den allgemeinen Marktanforderungen der Automobilindustrie nicht nur einmalig entsprechen, sondern sie müssen sich auch gegenüber der Produktions- und Nutzungsvariabilität bewähren. Die *Total Costs of Ownership* (TCO), zu Deutsch die *Gesamtkosten der Eigentümerschaft*, werden inzwischen als Schlüsselfunktion in der Qualität betrachtet, die es unter allen zulässigen Qualitätsstandards zu minimieren gilt.

Unter Einsatz von modellbasierter Simulation und statistischen Werkzeugen entwickelt diese Arbeit eine Methode um die versteckten Kosten zulässiger Qualitätsstandards und deren Erfüllung hinsichtlich existierender Anforderungen ohne wiederholtes Simulieren bewertbar zu machen. In diesem Zusammenhang wird ein modellbasierter *Produktions- und Lebensdauerorientierter Entwicklungsprozess* vorgestellt, wodurch ein Fahrzeughersteller beim zentralen TCO Minimierungsproblem in optimaler Weise unterstützt werden kann. Die stochastischen Einflussgrößen der Produktion und Nutzung werden unter Zuhilfenahme der *statistischen Prozesskontrolle* sowie mittels Umweltdaten- und Nutzungsprofilanalyse modelliert und in eine semi-physikalische Modellumgebung eingepflegt. Die zu berücksichtigenden Wahrscheinlichkeitsverteilungen werden dabei parametrisch oder nichtparametrisch aus vorliegenden Messdaten geschätzt. Zudem wird eine neue Versuchsanleitung, bestehend aus einer *Screening-Prozedur* und einer *Space-Filling-Prozedur*, präsentiert. Versuchspläne, wie das *Definitive Screening Design*, das *Latin Hypercube Design*, oder das in dieser Arbeit neu entwickelte *Depth Design* helfen dabei den zu Grunde liegenden hochdimensionalen Versuchsraum möglichst effizient untersuchbar zu machen. Die vorgeschlagene Versuchsanleitung wird um eine Methode erweitert, die eine Einbindung der Systemalterung unter Verwendung von Alterungsmodellen zulässt. Die zugrunde liegenden Verteilungen werden schließlich durch Einsatz von *Monte Carlo Methoden* und hochgenauen Metamodellen, wie dem *Multi-Layer Perceptron* oder *Radial Basis Function Networks*, repräsentativ abgedeckt, sodass versteckte Kosten zulässiger Qualitätsstandards und deren Erfüllung hinsichtlich existierender Anforderungen sofort bewertet werden können. Am Ende dieser Arbeit werden im Zuge einer Fallstudie die entwickelten Methoden am Beispiel eines Dieselmotors umgesetzt.

ABSTRACT

The powertrain development process of a vehicle continuously covers an increased number of complex systems for its major elements, the engine, the transmission, electric drives, the exhaust after treatment system and electronic controls. Here, the quality standards developed have to comply with the market requirements set out for the automotive industry and they need to be robust against the variability caused by production and lifetime usage. By now, the term *total costs of ownership* (TCO) is emerging more and more to the overall key function in quality that has to be minimized according to feasible quality standards. Under the usage of model-based simulation and statistics, this thesis establishes a method capable of instantly assessing the hidden costs of feasible quality standards and their compliance with the requirements given. In this regard, the *production and lifetime-oriented development* process is introduced so that a vehicle manufacturer can be optimally supported in solving the central TCO minimization task. In consideration of the *statistical process control* (SPC) and analysis procedures regarding environmental data and usage profiles, the stochastic factors incorporating the production and lifetime usage are modeled and implemented into a semi-physical simulation environment. The probability distributions necessary are estimated by parametric and nonparametric estimation approaches on basis of available measurement data. Moreover, an experimental guideline, consisting of a screening and a space-filling procedure, is given. The experimental designs applied, such as the *definitive screening design*, the *latin hypercube design* and our new development the *depth design* allow to efficiently explore the resulting high-dimensional feature space by computer simulations. The proposed guideline is then extended by a methodology on how to incorporate multiple aging states by aging models in order to capture the system deterioration of all factor combinations considered. The application of Monte Carlo methods on most accurate metamodels, such as the *multi-layer perceptron* or *radial basis function networks* eventually facilitates the coverage of the underlying feature space distributions in a representative manner. By this hidden costs of feasible quality standards and their compliance with the given requirements become instantly assessable. Finally, in a case study the methods developed are applied to a compression ignition engine.

Contents

List of Figures

V

List of Tables

Chapter 1

Preface

1.1 Acknowledgments

There are several persons who have largely contributed to the possibility of writing this doctoral thesis. I am honestly grateful to everyone of them. I express my deepest gratitude to my advisers Dr. Martin Schüßler and Univ.-Prof. Dipl.-Ing. Dr. Ernst Stadlober for their valuable support and all their scientific assistance over the last years. My special thanks go to DI Rolf Dreisbach, DI Markus Grubmüller and Dr. Theodor Sams for setting up a research project frame for this thesis. Moreover, I really want to thank my former lead engineer Dr. Ingo Allmer, who sadly deceased this year, for having supported me with his inexhaustible scientific knowledge since the very beginnings of my diploma thesis. Beside Dr. Martin Schüßler, I especially want to thank my colleague DI Marko Decker for showing me numerous application possibilities and for doing all the marketing work in the last year. I am grateful to all of my AVL colleagues for helping me to overcome all technical hurdles. DI Florian Ansperger, DI Peter Berger, Dr. Bernd Breitschädel, DI Thomas Glatz, DI Alexander Karlon, Dr. Michael Kordon, Dr. Christian Kozlik, Dr. Peter Lichtenberger, Dr. Waldemar Linares Rodriguez, DI Josef Macherhammer, Dr. Hannes Noll, Dr. Hannes Pucher, Dr. Markus Thaler, DI Hannes Wancura, DI Barbara Weirum, Dr. Michael Zallinger as well as Jakob Greisdorfer have to be explicitly mentioned at this point.

I dedicate this work to my family, my fiancée Elisabeth, my parents Mag. Markus and Marion Piffl, my brother DI Matthias Piffl and his wife Dr. Martina Piffl, my parents-in-law Michaela and Gerhard Zirngast, my sister-in-law Kristina Zirngast, my grandparents Gerhard, Regine, Meinhard and Annunziata and my uncle Lukas and aunt Angelika with their families. Without their unsurpassed backup I certainly wouldn't have reached the current state. Last but not least, I am also thankful to all my friends, who have enabled the necessary work life balance during the last years.

Ich danke Euch allen von ganzem Herzen.

1.2 Work Environment

The investigation of complex systems subject to many influence factors has been significantly affected by the new opportunities in the field of computer simulation software. Beside the functional range of software products available on the market, the high availability and broad acceptance in the industry are the main decision criteria.

AVL has decided on the well accepted commercial software packages MATLAB$^{\text{TM}}$ and Simulink$^{\text{TM}}$ in order to develop its powertrain and engine simulation tool MoBEO$^{\text{TM}}$, employed in this thesis. Both, Matlab and Simulink are offered by the MathWorks[1], Inc. company with headquarters in Natick, Massachusetts, USA. Matlab is a highly sophisticated mathematical programming language and interactive environment that is primarily consulted for data analysis, but also used to simulate and solve numerical tasks. Moreover, Matlab is the basis of the block diagram environment Simulink. Simulink is mainly used for time-controlled simulation and model-based design applications. The graphical editor, customizable block libraries and solvers make Simulink to a very well designed, technical intuitive and easy to work environment. Matlab and Simulink have been used to generate all the simulation data, analyzed and processed in the course of this thesis. Statistical computations and data analyzes have been performed with the Matlab toolboxes *neural network toolbox*, *optimization toolbox* and *statistics and machine learning toolbox*.

In addition to Matlab and Simulink, the software environment **R**$^{\text{TM}}$ has been employed during this thesis too. **R**, which origins at the university of New Zealand, is currently developed by the *R Development Core Team*[2]. It is an open source statistical programming language for data manipulation, calculation and graphical display. The effective data handling, the storage facility, the free access to a huge collection of functional packages for data analysis as well as the inclusion of conditionals, loops and user-defined recursive functions make **R** to the preferred choice among statisticians and data miners. In this thesis, the work environment **R** has been widely used to perform all statistical computations and to generate most of the figures (some figures were remodeled by Microsoft[3] **Excel**$^{\text{TM}}$ and **PowerPoint**$^{\text{TM}}$ 2013).

---

[1] www.mathworks.com
[2] cran.r-project.org
[3] www.mircosoft.com

Chapter 2

Introduction

Rinne and Mittag [1995] believe that quality, which can be seen as the level of satisfaction and the fulfillment of requirements, is directly linked to the success of a company. From their point of view, quality will directly lead, when present in an above-average form, to higher market shares, higher end-user confidence and so to higher revenues in the mid-term. Similarly to that, Montgomery [2012] describes quality as one of the most important decision factors in the selection among competing products and services. Understanding and manipulating quality in the right way leads to business success, growth and enhanced competitiveness.

Beside the quality requirements at the manufacturer's side and at the end-user's side, also the public society demands by law for particular quality standards. In the industrial sector, these legal requirements mainly request an environmentally-friendly product with regard to its production, usage and disposal. Consequently, it is not surprising that quality standards are continuously improved and assured by manufacturers. On this account, it is very important for AVL[1] customers (these are usually *Original Equipment Manufacturers* or OEMs) that the implemented quality standards comply with the overall requirements given and prevent the same from expensive hidden costs, that may arise from production or lifetime variability. In the automotive industry, the term *robustness investigation* has been established here for all investigations, accounting for the variability of the production and of the vehicle's usage so that hidden costs can be assessed and largely prevented. The term *Total Costs of Ownership* (TCO) emerges here more and more to the overall key function in quality that needs to be minimized under given quality constraints. In the past years, AVL has already performed robustness investigations that helped its customers to assess and refine their quality standards in terms of TCO. The parallel investigation of several identical test objects has allowed estimating the natural variation of crucial *response variables*, like fuel consumption or legally regulated emissions. Among others, this gave AVL the possibility to advise the OEMs in determining optimal hardware and software concepts, and in finding the engineering targets necessary. Nevertheless, so far many of these statements have been of qualitative nature and based on expert knowledge rather than on detailed mathematical calculations. Predicting TCO for a given quality standard requires, however, the availability of a closed mathematical method able to instantly provide quantitative statements about hidden costs and the

---

[1] www.avl.com

3

compliance with constraining requirements. On the basis of model-based engine optimization and the statistical concept of the *robust parameter design* this thesis will introduce a new approach, the so called *production and lifetime-oriented development* that allows the instant assessment of a quality standard in terms of given requirements and hidden costs. Once this is accomplished, TCO can be minimized over feasible quality standards.

2.1   Quality Management and the Minimization of the Total Costs of Ownership

Quality management is the fundamental basis of all principles and measures implemented by manufacturers in order to improve and assure quality. Due to Rinne and Mittag [1995] the concept of quality and its understanding have been subject to change over decades. While at the beginning quality considerations were mainly restricted to individual technical features, the general definition of quality expired soon and was expanded by the degree of suitability concerning intended purposes. In the automotive industry, this degree of suitability targets OEM requirements, legal requirements and end-user requirements. The most common requirements are listed in table 2.1 (cf. Timischl [1995]). As already pointed

OEM Requirements	Legal Requirements	End-User Expectations
Market acceptance	Emission regulations	Usability
Technical progress	Social purpose	Reliability
Cost effectiveness	Safety	Reasonable pricing
Faultless processes	Product liability	Delivery time
Risk assessment of liabilities and warranties		Appearance
Company image		Additional features

Table 2.1: Quality Requirements

out, the one-time compliance with given requirements is not sufficient anymore. By now, more importance is attached to quality standards that are, at least for a particular time period, insensitive to the variability of production processes and lifetime usage. Introduced robustness targets (e.g. low failure probabilities, limited penalty or warranty costs) shall guarantee that the product variability remains within specified limits so that hidden costs at the manufacturer's side, arising from legal penalties, end-user dissatisfaction, high warranty payments, or even vehicle call backs for instance, can be largely prevented. The transition to a higher quality standard is usually an expensive measure that needs to be balanced with all associated benefits for the manufacturer regarding the product's life cycle. Geissdörfer [2009] partitions the life cycle of a product into six stages reaching from the very first beginnings of the development stage to the point in time, when the product is eventually removed from the market. Figure 2.1 gives an overview of these stages and its associated costs drivers. Under the assumption of fixed profit margins the **L**ife **C**ycle **C**osts (LCCs) of the manufacturer directly determine the TCO of an end-user product considered. As outlined by Geissdörfer [2009], the economic literature interchangeably uses the terms **L**ife **C**ycle **C**osting (LCC) and TCO to describe these lifetime-related costs for both, the manufacturer and the end-user. This is, because TCO incorporates
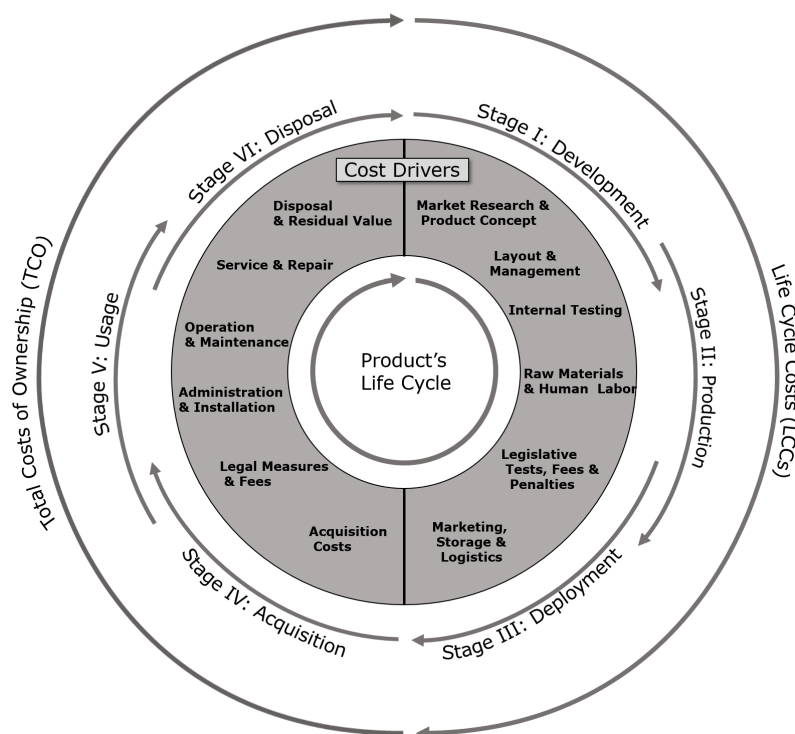
Figure 2.1: Total Costs of Ownership \\ Life Cycle Costing

the LCCs as acquisition costs. In this thesis the term TCO, and not LCC, is used to describe all costs of a product's life cycle. For Ellram [1995] TCO is a philosophy which is aimed at understanding the true costs of a particular product. Especially, in the word 'true', Götze and Weber [2008] recognize a reflection of the endeavor to completely assess all TCO. They infer that TCO denote more than a simple quantity, but an objective function to be minimized. After Ehrlenspiel et al. [2014] and Zehbold [1996], the variable part of TCO is thereby determined mainly by the quality provided by the manufacturer during the first stages of the product's life cycle. This statement is clearly illustrated by figure 2.2, originating these sources. Although the first phase, market research & product concept, amounts only small portion of the overall TCO incurred, it includes about 2/3 of the TCO determining factors. Moreover, the development and production stage include already more than 95% of the TCO determining factors. Hence, minimizing TCO can be almost entirely seen as a manufacturer's responsibility, and the provided quality during the development stage as the key figure to be manipulated. Consequently, the manufacturer should spend as much effort as technically feasible on the development stage. An automotive consulting firm may serve here as an engineering partner, advising its customers on optimally solving the central TCO optimization problem, captured in figure 2.3. There, the central task is to find a feasible quality standard or product specification that minimizes the TCO under the requirements given and the robustness targets set. Before investigating on how TCO response on a particular quality standard, it is worth to spend some lines on the quality management, covering all organizational measures intended to improve and assure quality standards at the manufacturer's side.
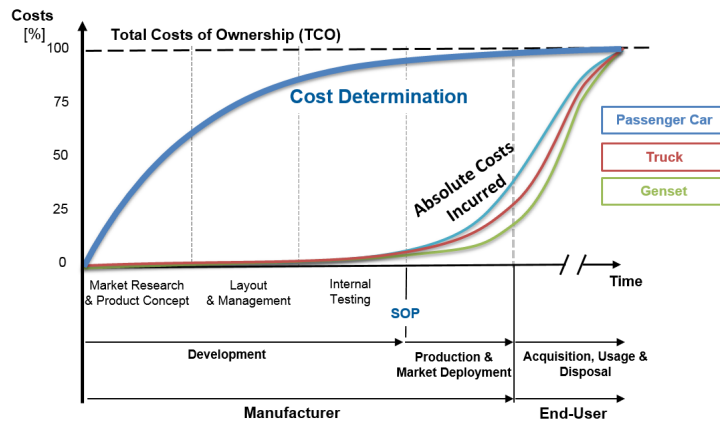
Figure 2.2: Determination of TCO
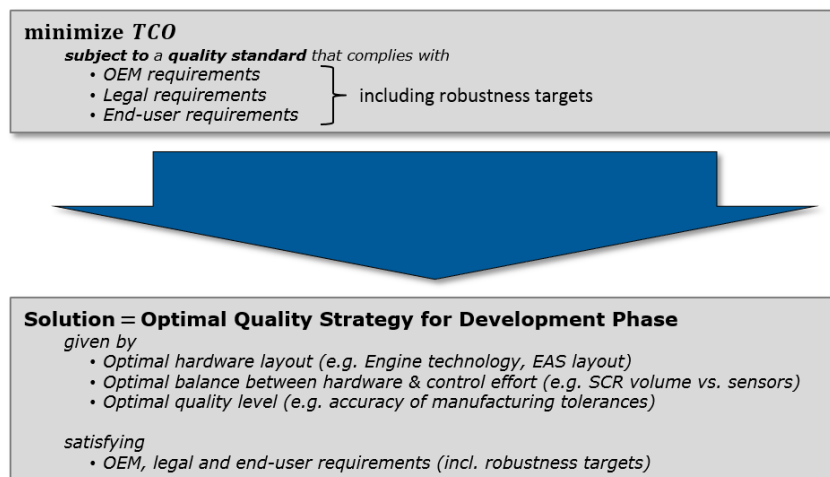
Figure 2.3: Minimization of TCO

2.1.1 Quality Improvement

Pindyck and Rubinfeld [2003] outline that a certain demand on the market is generally satisfied by several manufacturers. Basically, the manufacturers and end-users need to accept the existent acquisition prices on the market for the products offered. As described by Rinne and Mittag [1995], higher market shares are in the end the only possibility to achieve higher revenues for the manufacturer. If two comparable products have the same acquisition costs, it is expected that end-users tend to buy the more familiar product on the market. Hence, unless a manufacturer maintains a monopoly position, improving the quality standard is desirable, as it helps to increase the market shares and the company image in the mid-term. Moreover, a higher quality standard prevents the manufacturer from hidden costs that may arise due to unexpected product failures.

Dr. W. Edwards Deming (1900-1993) proposed to improve and assure quality by continuously going through the four stages of the *Plan*-*Do*-*Check*-*Act* (PDCA) cycle.

- **Plan.** The product to be manufactured needs to be planned before its serial production. This concerns the recognition of the current state and finding room for

improvements. The establishment of objectives in accordance with the deliverables allows the comparison between expected and achieved performance.

- **Do.** The product is manufactured by implementing the plan. Response variables are determined and corresponding data is collected in this phase.

- **Check.** Compare the actual results with the expected results that have been made during the planning phase. The deviation between actual and expected values needs to be assessed and converted into information.

- **Act.** Identify the root causes for the differences between actual and expected results, and either decrease or eliminate the deviations by corrective actions. The acquired knowledge can be applied in the next iteration during the planning phase.

As presented in figure 2.4, the successive iteration through the PDCA cycle shall improve the quality standard of a product. The PCDA cycle has its origin by a lecture of Deming
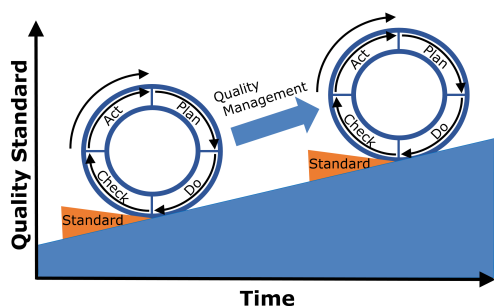
Figure 2.4: The Plan-Do-Check-Act Cycle
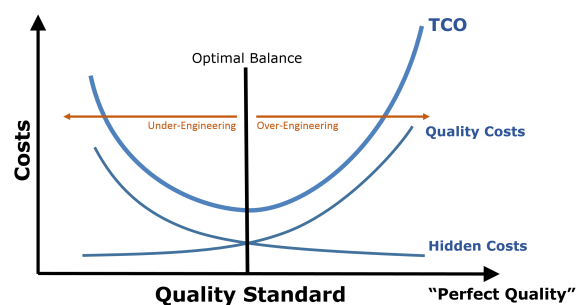
Figure 2.5: Balance between TCO and Quality

in Japan in 1950. His ideas found broad acceptance among the re-emerging Japanese companies after World War II. With Deming's approach these companies have been able to achieve high quality standards of Japanese products in the last century. Today, all modern quality management systems, just as robustness investigations, have their roots in the PDCA cycle.

The improvement of quality is usually an expensive measure, whose benefits are above all not directly noticeable for the manufacturer. Nevertheless, Timischl [1995] estimates the costs necessary to achieve a higher quality standard much lower than the hidden costs that would appear during the production and usage without the improvement measure. Yet, the achievement of "perfect quality" as such is not a desirable goal for the manufacturer. As described by Dong et al. [1994], the more the quality standard is improved, the higher become the "quality costs" and the lower become the benefits in terms of hidden costs. Finally, there exists a point where quality costs overshoot the savings in terms of hidden costs, as illustrated in figure 2.5. That quality standard may be denoted as "over-engineered" system. While over-engineered systems are generally accepted in practice, the counterpart, that is an "under-engineered" system has to be avoided at all costs. Dobler et al. [2003] believe that achieving a balance between quality and hidden costs is the true challenge that must be overcome to minimize TCO. And, in fact, quality costs

can be appraised by the OEM in the development stage, while the associated hidden costs are often unknown.

Today, quality management is playing such an important role that large-scale customers urge manufacturers heavily to certify their quality management systems by independent service companies. In the automotive industry the certification of all measures, which assure quality, is mandatory. Beside service companies as the **A**merican **S**ociety for **Q**uality (ASQ)[2] or the **E**uropean **O**rganization for **Q**uality (EOQ)[3] the most notable is the **I**nternational **O**rganization for **S**tandardization (ISO)[4], with headquarters in Geneva, Switzerland.

2.1.2 Quality Assurance

Improving quality requires the continuous assurance of quality standards already achieved. Quality assurance refers to all activities that ensure the compliance of existent requirements. Due to company-internal peculiarities, quality assurance measures have been significantly differed from one to the other company. In 1987, the *International Organization for Standardization* introduced the norm series ISO 9000 that established the first standardization in quality management. ISO 9000 defines, provides and explains effective quality assurance measures that prevent from defective products and poor service performance.

The code numbers ISO 9001 and ISO 16949 are essential for the automotive industry. ISO 9001 includes a multiplicity of minimum guidelines, whereof all concern the quality management and are a manual on how products satisfy the requirements of end-users at best. ISO 16949 is part of the ISO 9001 family, which regulates the development, the production, the assembly and all kind of services in the automotive industry.

It is the increasing complexity of the offered products on the market that cause quality assurance procedures to rise continuously in costs and time, making the supervision of quality during the production chain no longer sustainable. Especially in the automotive industry, ISO has become an inevitable standard that sustains competitiveness of manufacturers and service providers all over the globe. Today, ISO is the market leader in terms of quality assurance and the world's largest developer of international quality management standards.

2.2 Production and Lifetime-Oriented Development

ISO provides also a generally accepted definition of quality. After ISO 9000:2005 quality refers to *the degree to which a set of inherent characteristics fulfills requirements* (cf. table 2.1). While AVL has been supporting its customers in achieving quality related requirements by concept or by feasibility studies for years, lately the AVL support also copes with robustness investigations, giving the possibility to better assess the overall TCO. Nevertheless, the proposals given have been often based on expert knowledge. Therefore,

---

[2]www.asq.org
[3]www.eoq.org
[4]www.iso.org

an unprecedented solution of the central TCO minimization task in figure 2.3 requires the ability to quantitatively assess a quality standard in terms of requirements and hidden costs.

The TCO correspond to a function in quality that is provided by the manufacturer at the development stage (cf. figure 2.2). Two objectives have to be met in order to find an optimal quality standard in terms of the central TCO minimization task. On the one hand, the relationship between a quality standard provided and the hidden portion of the TCO involved, such as vehicle call-backs, warranty services or legal penalties, have to be revealed. On the other hand, the compliance of the quality standard provided needs to be measurable in terms of the requirements of table 2.1. The explanations of section 2.1 make clear that TCO is mainly determined by *controllable factors* set in the development stage (cf. figure 2.2). Eventually, the *uncontrollable factors* of the production and usage stage determine the actual amount of TCO, and they put an implemented quality standard to the test under the requirements given. The task of making the correct decisions regarding controllable factors that are assessed in terms of uncontrollable factors gives rise to the **R**obust **P**arameter **D**esign (RPD) approach, firstly developed by the Japanese engineer Genichi Taguchi (cf. Taguchi and Wu [1980] and Taguchi [1987]).

As Montgomery [2013] writes, *the RPD is an approach to product realization activities that focuses on correctly setting controllable factors of a process (or system) to achieve two objectives: (1) to ensure that the mean of the output response is at a desired target, and (2) to ensure that the variability around this target value (caused by uncontrollable factors) is as small as possible.* Figure 2.6 illustrates the RPD approach. The process
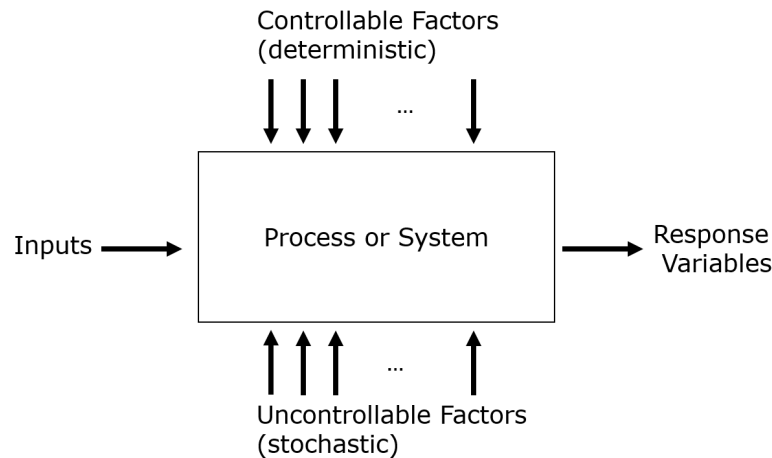
Figure 2.6: Robust Parameter Design

basically denotes a combination of resources, like methods, machines or human work, transforming an *input* into an *output*. With regard to the automotive industry, the input usually denotes an operable system, such as a complete vehicle, a stand-alone engine or on an **E**xhaust **A**ftertreatment **S**ystem (EAS). In this thesis it denotes for simplicity an engine environment consisting of the engine, the EAS, the **E**ngine **C**ontrol **U**nit (ECU) as well as all actuators and sensors required by the ECU. Thus, the process can be considered here as the experimental procedure, where the engine environment is exposed to particular settings of the controllable and uncontrollable factors. The controllable factors

are the variable parameters of a quality standard that is to be realized for the objective engine environment. As the name suggests, controllable factors can be specifically set. Subsequently, we refer to them as *open system parameters*. Montgomery [2013] defines the uncontrollable factors as parameters that define manufacturing tolerances (e.g. dimensions of catalysts), different usage behavior (e.g. usage profile: 40% city driving, 50% highway driving, 10% rural driving) and various environmental conditions (e.g. ambient temperature, pressure and humidity). Uncontrollable factors are of stochastic nature and are linked to probability distributions evidently. The output refers to one or multiple data recording channels, referred to as *response variables*. The response variables could be the consumption of operating fluids, all kinds of emissions, or the durability of individual engine components. Consequently, the response variables allow to meet the two objectives for the central TCO minimization task, as the hidden portion of TCO and the compliance with general requirements becomes assessable.

Summing up, resolving the central TCO minimization task requires including the uncontrollable factors of the production and lifetime usage into the development process of a vehicle. Particularly for the automotive industry, we denote this development approach as the *production and lifetime-oriented development*. This denomination substantiates the joint consideration of the development and usage stage respectively. A sketch of the new approach is given by figure 2.7. On the one side, there are the factors of the de-
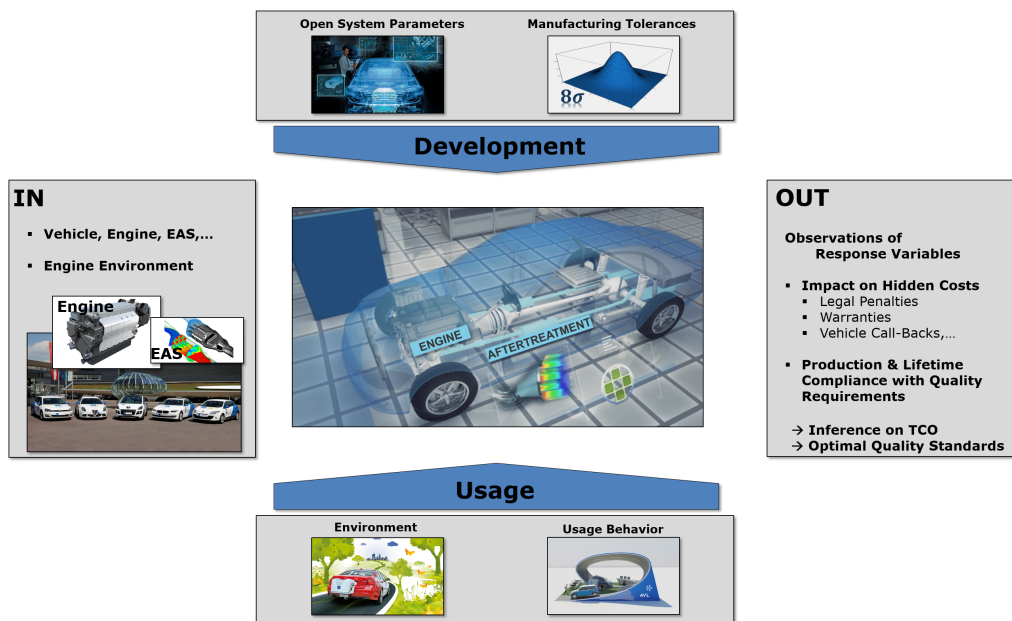
Figure 2.7: Production and Lifetime-Oriented Development

velopment stage. They consist of all open system parameters and a specification table for manufacturing tolerances. Parameters regarding the environmental conditions and usage behavior review the investigated quality standard. They refer to the usage stage, positioned on the opposite side. Based on the RPD approach, a remarkable extension of the new development approach is that manufacturing tolerances are passively controllable (nevertheless stochastic factors) by open system parameters. This is because the OEM can generally specify the tolerance limits of the obstructed vehicle components by

a change of the production quality, a transition to another hardware strategy, or a switch of the suppliers. Although the development and usage stage are considered at once, an independent treatment of both stages shall be assumed. While ambient temperature may depend on ambient humidity, and manufacturing tolerances on open system parameters, the factors of the usage stage are assumed to be independent of the factors of the development stage.

Launching the production and lifetime-oriented development process with a real engine environment proves to be not feasible in practice. This is because the investigated factor combinations must achieve a representative image of the stochastic factors included. In other words, the investigated quality standards need to be completely assessed under the variability of production and lifetime usage. The representative coverage becomes, in fact, the more challenging, the more stochastic factors are included. Moreover, the representation of a manufacturing specification table is almost not feasible with a real engine environment. The engine would have to be completely reassembled with differently specified components for each factor combination considered. Although this drawback does not necessarily apply to open system parameters, environmental conditions and usage behavior (they can be fairly precisely simulated on a test bed), the time and financial effort for the experimental setup and subsequent tests would be disproportionally high. Here, AVL overcomes these problems by the semi-physical simulation tool MoBEO$^{\text{TM}}$, capable of accurately predicting all kind of processes taking place inside the engine environment. MoBEO provides the outstanding advantage that different factor combinations concerning the development and usage stage are easily implemented and simulated faster than real time.

2.3 Model-Based Development

Since 2008, AVL has been developing the MATLAB$^{\text{TM}}$ Simulink$^{\text{TM}}$ based simulation tool MoBEO (cf. Schüßler et al. [2008]). MoBEO, which stands for ***Mo**del **B**ased **E**ngine **O**ptimization*, is a model-based development approach to powertrain modeling that covers transient thermodynamics and tail pipe emissions.

The semi-physical tool MoBEO combines fast calculating empirical models with physical components in order to increase the range of application. Moreover, the associated real time capable engine and EAS simulations support the whole development process of a vehicle in the fields of design, engine control software development and calibration. The prediction accuracy of MoBEO generally ranges from approximately 3% for fuel consumption, over $10\,K$ ($20\,K$) for temperatures at the intake side (at the exhaust side) to 15% for tailpipe emissions. The high prediction accuracy provides the advantage to switch from real to virtual engine and exhaust aftertreatment testing. There, different hardware and software concepts can be investigated under specified stationary mappings or time-resolved input traces. In general, these input traces are given by operational points, consisting of engine speed, engine torque and environmental conditions. If a stand-alone exhaust aftertreatment system is simulated on its own, the required simulation input corresponds to a mapping or time series of temperatures, pressures, mass flows and gas compositions. MoBEO also allows the application of aging models so that engine deterioration and exhaust aftertreatment performance losses can be simulated. For all its

advantages, MoBEO is a suitable tool to conduct the introduced production and lifetime-oriented development process.

MoBEO typically provides a model engine environment, which consists of four main interacting layers that are connected by data buses. One layer contains all sensors, another all actuators, another the combustion engine and the EAS components and another the *Engine Control Unit* (ECU). These layers again consist of sublayers, modeling hardware components or special ECU controls. The most accurate model prediction of response variables demands for setting all variable model parameters in an optimal way. As highlighted by figure 2.8, the real engine environment can be either simulated by the *Hardware in the Loop* (HiL) or the *Model in the Loop* (MiL) approach. With regard to the HiL
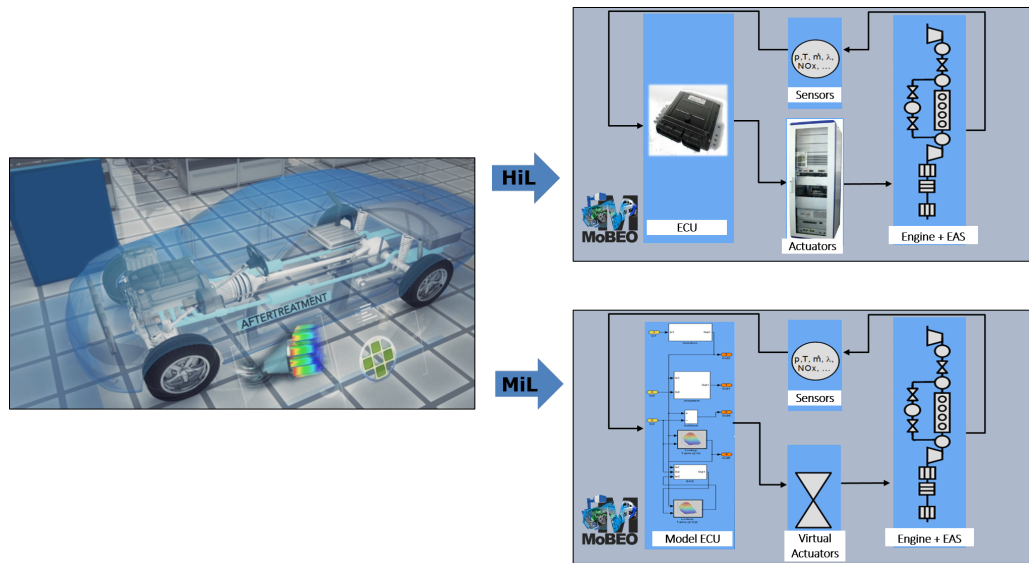
Figure 2.8: Model-Based Engine Optimization

approach, a real ECU is linked to the model environment, making allowance for ECU software as well as data set validation or precalibration of all ECU functions. The most realistic ECU behavior is achieved, if real actuators, like valves or injectors, are additionally installed. The HiL simulator has, however, the drawback that simulations are not faster-than-real-time. Furthermore, the open system parameters may concern the installed hardware so that expensive multiple implementation effort might be necessary. This specific drawback of the HiL approach can be resolved by the MiL approach. The MiL approach provides a software ECU and virtual actuators, which enable faster-than-real-time simulations on the one hand, and the possibility of easily modifying actuators and ECU on the other hand. Indeed, the consideration of a few basic ECU functions proves often to be sufficient so that the calibration effort necessary remains within a reasonable time frame.

The semi-physical simulation tool MoBEO exemplarily overcomes the problems that relate to the conduction of the introduced production and lifetime-oriented process with a real engine environment. As outlined in this section, MoBEO can cheaply, accurately and quickly investigate arbitrary factor combinations. Although MoBEO is an integral part for transforming quality standards in observations of relevant response variables,

simulation time and costs are usually limited. In general, a representative consideration of the stochastic factors is however not ensured. Thus, the variability of the production and lifetime usage has to be reflected by a limited number of experiments. Once this is accomplished, it follows from the RPD approach that the associated observations are representative for the distributions of the response variables.

The problem of inferencing from a limited number of experiments on probability distributions directly refers to the field of statistics. Indeed, statistical tools allow generating random variates of these distributions, if and only if the assumptions of the stochastic factors are adequately fulfilled and efficiently as well as accurately processed. It is the central target of this thesis to achieve these requirements and to generate samples of the required probability distributions. Achieving this target means in detail addressing the following four work items.

1. *Probabilistic Modeling & Implementation*,
2. *Design of Experiments*,
3. *Lifetime Simulations*,
4. *Metamodeling Approaches and **M**onte **C**arlo (MC) Simulations*.

As illustrated by figure 2.9, these items are the basis of a production and lifetime-oriented development process that is technically feasible. Each of the listed work items will be discussed in the subsequent chapters of this thesis. Chapter 3 focuses on the careful probabilistic modeling of the stochastic factors. The corresponding implementation of these factors into MoBEO is discussed in that chapter too. The second item, "*Design of Experiments*", is explained in chapter 4. Strategies on how to optimally generate factor combinations are presented there. Chapter 5 gives rise to the simulation procedure, conducted with MoBEO, and provides a methodology for considering one or multiple aging states. Chapter 6 presents state-of-the-art regression model approaches that can establish a functional relationship between the factors involved and response variables considered. These models are capable of predicting several thousand of factor combinations within seconds. The application of MC simulation procedures yields observations that are representative for the distributions of the response variables. Chapter 7 concludes this thesis with a case study, where the distributions found are converted into quantitative statements regarding the given requirements and in terms of hidden costs.
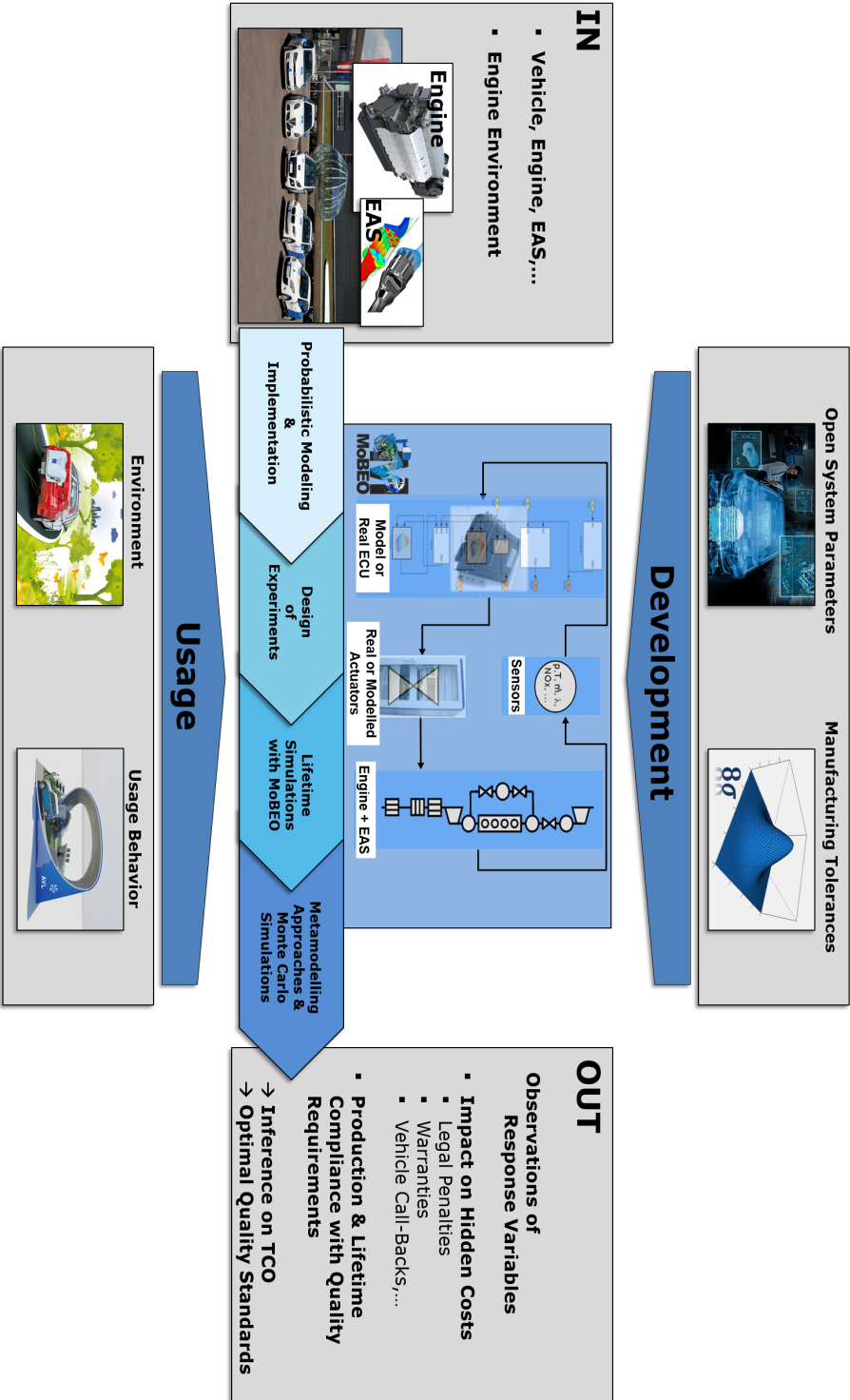
Figure 2.9: Launching the Production and Lifetime-Oriented Development under MoBEO and Statistics

Chapter 3

Probabilistic Modeling and Implementation

The introduced production and lifetime-oriented development process is intended to evaluate quality standards in terms of the variability of production and lifetime usage. This characteristic lays the exclusive foundation for determining a solution of the central minimization task regarding TCO. Although the simulation of different factor combinations is overcome with the semi-physical simulation tool MoBEO, the production and lifetime usage cannot be simulated in a representative manner. This means that the simulation capacity of MoBEO is generally not sufficient to represent the distributions of the stochastic factors included. Before facing the problem of a representative coverage, it is however important to clarify how these distributions are derived, modeled and implemented into the process. On this account, section 3.1 firstly deals with the factors of the development stage. Especially, the probabilistic treatment of manufacturing tolerances is discussed in the relating subsections. In contrast, section 3.2 copes with the stochastic factors of the usage stage. The final section 3.3 of this chapter defines the feature space, that is the set of all factor combinations.

3.1 Modeling and Implementing the Development Stage

Usually, the development stage is the period where automotive consulting firms support the OEMs in terms of quality-related decisions. The semi-physical simulation tool MoBEO shall serve here as a virtual test-bed, where the factors regarding the development stage, that are open system parameters and manufacturing tolerances, shall be varied and simulated. Before starting the first simulation, it has to be clarified which engine environment is the natural starting point, and what deviation measures are realistic.

The investigated factors basically refer to discrete or continuous features. On the one hand, discrete features can only take finitely many values or *levels*, and they do not necessarily need to be numeric (cf. Fahrmeir et al. [2010]). The substrate material of an EAS monolith or different control strategies concerning the air path relate to discrete features for example. Their factor levels are commonly referred as to {"0","1","2","3",...}, es-

pecially when more than three factor levels are considered (cf. Montgomery [2013]). On the other hand, continuous features can realize on infinite possible values. The sensor position on a pipe, the volume of a catalyst or calibration parameters set in the ECU are typical examples for continuous features. In contrast to discrete features, which are exclusively modeled by qualitative factors, continuous features can be considered by both qualitative and quantitative factors. In case of a qualitative consideration, the levels are often denoted by $\{" - "," + "\}$. When there is also a center value available, one uses $\{" - 1","0"," + 1"\}$ too. The consideration of quantitative factors is especially practicable for computer experiments, where arbitrary values can be easily implemented between the interval $[-1, +1]$.

In the introduced production and lifetime-oriented development process all involved features can be assumed as random variables. As a consequence of the underlying RPD approach, the observed response variables, which refer to discrete or continuous features (e.g. fuel consumption, the light up of the **M**alfunction **I**ndicator **L**ight (MIL), etc.), denote random variables $Y_1 \sim F_{Y_1}, \ Y_2 \sim F_{Y_2}, \ldots, \ Y_r \sim F_{Y_r}$. Given that open system parameters are independent and not subject to any company internal preferences, they can be assumed for reasons of consistency as independent uniformly distributed random variables. These are considered by discrete or continuous features $U_1, \ldots, U_{\tilde{k}}$. It follows from the RPD approach that the associated realizations $u_1, \ldots, u_{\tilde{k}}$ determine the mean values μ_{Y_s} of the considered response variables for $s = 1, \ldots, r$. The variability of the response variables is determined by the factors regarding manufacturing tolerances. In the automotive industry, manufacturing tolerances refer frequently to continuous features, just as the cell densities or volumes of catalysts. In this thesis, factors regarding manufacturing tolerances denote continuous random variables $X_1, \ldots, X_{\check{k}}$, which transmit their variability on the response variables. Making statements on how $Y_1, \ldots, Y_r$ respond to different combinations of development decisions and manufacturing tolerances requires appropriate data to be gathered.

The steady implementation and simulation of the development stage is overcome by the simulation tool MoBEO. Open system parameters and manufacturing tolerances are implemented by altering the existing layers, their interaction, the included functions or the parameters loaded. Basically, it can be said that the consideration of a discrete random variable $U_{\tilde{j}}$, which represents different boost pressure concepts for example, comes along with more modification effort, because layers, wiring, functions and parameters need to be simultaneously modified. The implementation of a continuous random variable $U_{\tilde{j}}$ is realized, in contrast, by parameters only. Because of the higher implementation effort in terms of discrete random variables, the number of continuous random variables considered is usually larger. Especially, the careful probabilistic handling and implementation of manufacturing tolerances plays a central role for implementing the development stage. This is because the number of considerable manufacturing tolerances $\check{k}$ far exceeds the number of simultaneously considered open system parameters $\tilde{k}$. Additionally, the distribution structure regarding manufacturing tolerances is more complex. While $U_1, \ldots, U_{\tilde{k}}$ are considered as independent uniformly distributed random variables, $X_1, \ldots, X_{\check{k}}$ are not necessarily independent random variables and may follow other than uniform distributions, as described by the **S**tatistical **P**rocess **C**ontrol (SPC).

3.1.1   Statistical Process Control

The SPC is regulated by ISO and origins the *Six Sigma*[1] approach, firstly published by
Motorola in the mid 80s. SPC provides the probabilistic background for manufacturing
tolerances. According to Fischer [2011] the manufacturing tolerance of a produced com-
ponent is determined by physical quantities aiming hardware related features that deviate
from their nominal design values. These features are subject, as applicable, to deviations
of the form, size, orientation or location.

The actual causes of manufacturing tolerances are versatile. Hochmuth et al. [2007] clas-
sifies all causes according to the *four M's*, given by *Material, Machine, Method, Men*. Raw
material used for manufacturing is not completely identical. Different winning processes,
impurities of the resources and transit under different conditions cause a natural devia-
tion of the quality. Regular wear and tear of the construction unit makes the permanent
output of identical products impossible too. Beside the fact that production methods
may differ from one to another factory, also outer reasons do sometimes force engineers
to instantly change scheduled production processes. The list is completed with the incon-
stancy of human labor, which has a substantial impact on almost all production processes
worldwide. The exhibition and adherence of manufacturing tolerances can be seen as a
major pillar of the engineer's responsibility.

Manufacturing tolerances are controlled for two important reasons. At first, development
procedures, conducted on a little number of prototypes, should remain valid and appli-
cable for the whole subsequent production series. The second reason is that components,
obstructed inside the powertrain, must not significantly endanger its performance and
need to get easily exchanged in a fault event (cf. Dobler et al. [2003]). At the beginnings
of quality assurance, quality of individual features was simply ensured by monitoring all
produced components. If any key features did not correspond to the expectations made,
the relating components were sorted out before delivery. Eventually in the 1930s, She-
wart [1939] proposed statistical sampling and assessment methods in order to intervene
defective processes. SPC, as regulated by ISO 11462:2001, combines all statistical sam-
pling methods to prevent from unacceptable manufacturing tolerances during production.
Moreover, it provides all measures to monitor, evaluate and adjust running processes. The
basic principle of SPC is the statistical inference to the whole product series on the basis
of random samples that need to be drawn out of the production line. As Montgomery
[2013] says, "*a random sample is a sample that has been selected from the population in
such a way that every possible sample has an equal probability of being selected*". Depen-
dent on the scale of the monitored feature, *counting methods* or *measuring methods* are
applied to make these statistical assessments. ISO 2859:2006 and ISO 3951:2013 provide
detailed information concerning counting methods and measuring methods, respectively.
Unless the monitored feature is not qualitative, measuring methods are usually preferred
to counting methods, because they contain more information.

Assuming that all features observed are quantitative, manufacturing tolerances refer to the
continuous random variables $X_1, \ldots, X_{\breve{k}}$, whose realizations $x_1, \ldots, x_{\breve{k}}$ describe to which
extend nominal design values $m_1, \ldots, m_{\breve{k}}$ have been met. Measuring methods assume
that the influence of the four M-causes balances on average for each monitored feature

---

[1]cf. Rinne and Mittag [1995], Timischl [1995], Brussee [2012] or George et al. [2004]

$X_{\breve{j}}$ (cf. Mischke [1980]). It is moreover assumed that extensive deviations from nominal values result of unusual production incidents. This concept corresponds to a well known mathematical statement, the so called **C**entral **L**imit **T**heorem (CLT) (cf. Fahrmeir et al. [2010]). Thus, measuring methods assume all features to be independent random variables that are subject to the normal distribution. This is formally expressed by the notation $X_{\breve{j}} \sim N\left(\mu_{\breve{j}}, \sigma_{\breve{j}}^2\right)$ for $\breve{j} = 1, \ldots, \breve{k}$. Although the bell-shaped density of the normal distribution is probably well known, the functional form should be reminded at this point.

Definition 3.1. (*Density of the Normal Distribution*)

*Let $X \sim N(\mu, \sigma^2)$ be a normally distributed random variable. Then, the **p**robability **d**ensity **f**unction (pdf) of X corresponds to*

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \tag{3.1}$$

where

$$x \in \mathbb{R}, \mathbb{E}(X) = \mu \ and \ VAR(X) = \sigma^2 > 0.$$

As apparent from the definition above, the support of the normal distribution's density is the complete closed real number scale, which theoretically allows arbitrarily large deviations from a given nominal design value. It seems intuitively clear that there must exist tolerance limits that ensure the capability of the components produced. Here, SPC defines production capability by the tolerance interval $T = [LSL, USL]$, wherein a monitored feature X must realize with a minimum amount of probability. Thereby, LSL denotes the **L**ower **S**pecification **L**imit, and USL denotes the **U**pper **S**pecification **L**imit of feature X.

Definition 3.2. (*Process Capability*)

In the automotive industry, a production process is usually denoted as capable, if the monitored features $X_{\breve{j}} \sim N(\mu_{\breve{j}}, \sigma_{\breve{j}})$ exhibit mean values $\mu_{\breve{j}}$ that correspond approximately to the nominal values $m_{\breve{j}}$, and if the process capability value

$$C_{p_{\breve{j}}} = \frac{USL_{\breve{j}} - LSL_{\breve{j}}}{6 \cdot \sigma_{\breve{j}}} \ with \ -\infty \le LSL_{\breve{j}} < \mu_{\breve{j}} < USL_{\breve{j}} \le \infty \tag{3.2}$$

exceeds $4/3$ for all $\breve{j} = 1, \ldots, \breve{k}$. In other words, a production process is capable, if

$$\forall \breve{j} : LSL_{\breve{j}} \overset{!}{\le} \mu_{\breve{j}} - 4 \cdot \sigma_{\breve{j}} \ and \ USL_{\breve{j}} \overset{!}{\ge} \mu_{\breve{j}} + 4 \cdot \sigma_{\breve{j}} \ and \ \mu_{\breve{j}} \approx m_{\breve{j}}. \tag{3.3}$$

The normal distribution $N(\mu, \sigma^2)$ possesses 99.994% probability mass below the 8σ-tolerance interval $T = [\mu - 4 \cdot \sigma, \mu + 4 \cdot \sigma]$. This means that less than 7 of 100.000 produced components are defective with regard to every monitored feature $X_{\breve{j}}$ (cf. 1st distribution in figure 3.1). If one feature $X_{\breve{j}}$ is exclusively monitored for a produced component, the associated distribution is said to completely describe the underlying production process. According to Timischl [1995], SPC distinguishes two different causes for unsatisfactory production quality. These are *incapable* and *off-center processes*.

Definition 3.3. (*Incapable Production Process*)

An incapable process provides at least a feature $X_{\breve{j}}$ that exhibits a process variance $\sigma_{\breve{j}}^2$ that is too large compared with the outlined specification limits (cf. 2nd distribution in figure 3.1). Hence,

$$\exists \breve{j} : LSL_{\breve{j}} > \mu_{\breve{j}} - 4 \cdot \sigma_{\breve{j}} \ and \ USL_{\breve{j}} < \mu_{\breve{j}} + 4 \cdot \sigma_{\breve{j}}. \tag{3.4}$$

Natural wear and tear of the construction unit may cause one or more $\mu_{\breve{j}}$ to drift away from $m_{\breve{j}}$ towards one of the corresponding specification limits.

Definition 3.4. (*Off-Center Production Process*)

An off-center process is characterized by at least one process mean that is too far away from the corresponding nominal value (cf. 3rd distribution in figure 3.1), i.e.

$$\exists \breve{j} : \left(LSL_{\breve{j}} > \mu_{\breve{j}} - 4 \cdot \sigma_{\breve{j}} \ or \ USL_{\breve{j}} < \mu_{\breve{j}} + 4 \cdot \sigma_{\breve{j}} \right) \ and \ \mu_{\breve{j}} \neq m_{\breve{j}}. \tag{3.5}$$
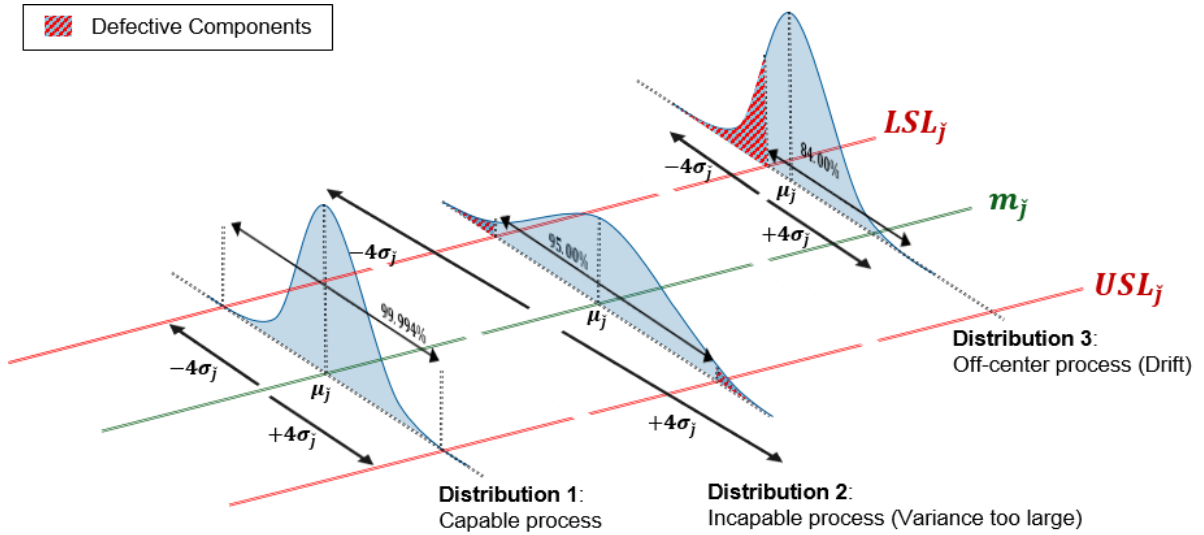
Figure 3.1: Process Capability

Drifting of the process mean is not considered by the process capability value. In general, features $X_{\breve{j}}$ need to be considered by a more deterministic process capability value $C_{p\kappa_{\breve{j}}}$.

Definition 3.5. (*Capable Production Process Considering Process Drifting*)

A production process is denoted as capable and safe from process drifting, if the process capability values satisfy

$$\forall \breve{j} : C_{p\kappa_{\breve{j}}} = C_{p_{\breve{j}}} \cdot \left(1 - \kappa_{\breve{j}} \right) \ where \ \kappa_{\breve{j}} = \frac{|m_{\breve{j}} - \mu_{\breve{j}}|}{T_{\breve{j}}/2} \leq 1. \tag{3.6}$$

This process capability value additionally considers, beside the process variance, the process mean $\mu_{\breve{j}}$ within the tolerance interval $T_{\breve{j}}$. Given that the process mean $\mu_{\breve{j}}$ of $X_{\breve{j}}$

corresponds to the nominal value $\mu_{\breve{j}}$ (i.e. $\mu_{\breve{j}} = m_{\breve{j}}$), $\kappa_{\breve{j}} = 0$ and $C_{p_{\breve{j}}} = C_{p\kappa_{\breve{j}}}$ is implied. For all other cases, the $C_{p\kappa_{\breve{j}}}$-value is more restrictive than the $C_{p_{\breve{j}}}$-value so that $C_{p\kappa_{\breve{j}}} \leq C_{p_{\breve{j}}}$ follows for all $\breve{j}$. Process drifts of $\mu_{\breve{j}}$ from $m_{\breve{j}}$ up to $2\sigma_{\breve{j}}$ away are assumed to be realistic in engineering (i.e. $|m_{\breve{j}} - \mu_{\breve{j}}| \geq 2 \cdot \sigma_{\breve{j}}$). If the outlined specification limits of a feature $X_{\breve{j}}$ span at least the tolerance interval $T_{\breve{j}} = \left[\mu_{\breve{j}} - 6 \cdot \sigma_{\breve{j}}, \mu_{\breve{j}} + 6 \cdot \sigma_{\breve{j}} \right]$, it follows that

$$C_{p\kappa_{\breve{j}}} = C_{p_{\breve{j}}} \cdot \left(1 - \kappa_{\breve{j}}\right) = \frac{T_{\breve{j}}}{6 \cdot \sigma} \cdot \left(1 - \frac{|m_{\breve{j}} - \mu_{\breve{j}}|}{T_{\breve{j}}/2}\right) \geq \frac{12 \cdot \sigma_{\breve{j}}}{6 \cdot \sigma_{\breve{j}}} \cdot \left(1 - \frac{2 \cdot \sigma_{\breve{j}}}{6 \cdot \sigma_{\breve{j}}}\right) = 2 \cdot \frac{2}{3} = 4/3. \quad (3.7)$$

$C_{p\kappa_{\breve{j}}} \geq 4/3$ implies that approximately 99.9999998% of feature $X_{\breve{j}}$ realizes in $T_{\breve{j}}$ (cf. figure 3.2). Consequently, a process is capable and safe from process drifting, when it is outlined by a 12σ-tolerance interval. Although the 12σ-tolerance interval is the basis of Six Sigma strategies, the fulfillment of $C_{p\kappa} = 4/3$ is rather of theoretical concern (cf. Toutenburg and Knöfel [2008]). In the context of this work, process capability will be assumed without process drifting and with a capability-process-value that equals $C_p = 4/3$ for all normally distributed features. Consequently, unless specified differently, available tolerance specification limits can be assumed to span 8σ-tolerance intervals.

Most commonly, manufacturing tolerances in terms of a feature $X_{\breve{j}}$ are outlined by
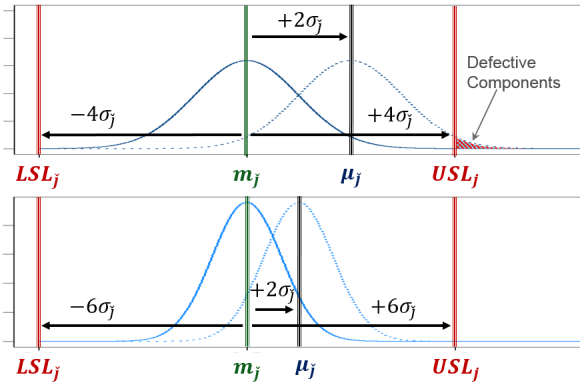
Figure 3.2: 12σ: Off-Center Processes
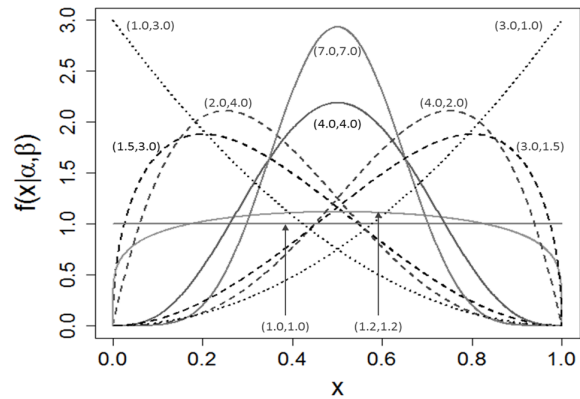
Figure 3.3: Densities of Different Beta Distributions

the specification limits $LSL_{\breve{j}}$ and $USL_{\breve{j}}$ only. In fact, it is the distribution of $X_{\breve{j}}$ that completely describes the issue of manufacturing tolerances. If no further distribution information is supplied and both provided limits have (approximately) the same distance from the nominal design value $m_{\breve{j}}$, the application of the normal distribution is suggested. Assuming process capability, the parameters $\mu_{\breve{j}}$ and $\sigma_{\breve{j}}^2$ of the normal distribution are consequently determined by

$$\mu_{\breve{j}} = m_{\breve{j}} \text{ and } \sigma_{\breve{j}}^2 = \left(\frac{USL_{\breve{j}} - LSL_{\breve{j}}}{8}\right)^2. \quad (3.8)$$

Although SPC is strongly linked to the normal distribution, monitored features do sometimes realize in a uniform or skewed manner. Furthermore, manufacturers may still monitor the whole production and sort out defective components. For that reason, features

$X_{\tilde{j}}$ shall be generally allowed to follow other than normal distributions. This shall be emphasized by the notation $X_{\tilde{j}} \sim F_{X_{\tilde{j}}}$ (or $X_{\tilde{j}} \sim F_{\tilde{j}}$) in this thesis. The uniform distribution, for example, seems a good choice for all features that realize uniformly within their specification limits. The *Lognormal distribution*, the *chi-squared distribution* or the Beta distribution cope among others with skewed deviation behavior. Especially, the Beta distribution provides useful properties in the context of manufacturing tolerances, as highlighted by Lin et al. [1997]. Various density shapes can be achieved by setting only two parameters α and β. The Beta distribution also eliminates the explicit need of the uniform distribution with parameters $\alpha = 1$ and $\beta = 1$. Moreover, the Beta distribution is able to represent the normal distribution over a finite interval, making allowance for sort out procedures in the forefront of the delivery. The various shapes of different *pdf*s of possible Beta distributions are illustrated in figure 3.3. The *pdf* of the Beta distribution is usually defined over the unit interval $[0, 1]$ as follows (cf. Hahn and Shapiro [1967]),

Definition 3.6. (*Density of the Beta Distribution*)

Let $X \sim Beta(\alpha, \beta)$ denote a Beta distributed random variable. Then, the pdf of X is given by

$$f(x \,|\, \alpha, \beta) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)}, \tag{3.9}$$

where

$$0 \le x \le 1 \ \ and \ \alpha, \beta > 0.$$

The denominator $B(\alpha, \beta) = \int_{0}^{1} t^{\alpha-1} (1-t)^{\beta-1} dt$ denotes the complete Beta function.

In principle, the definition of the Beta distribution can be generalized to an arbitrary interval $[a, b] \subset \mathbb{R}$ (cf. Casella and Berger [2006]). However, as alluded in subsection 4.1.1 the standardization, for example to the domain $[0, 1]$, is often a meaningful preparatory action before starting with a statistical evaluation of several features.
Applying the Beta distribution requires generalizing the definition of process capability. Before that, the cumulative distribution function and the theoretical quantile function must be defined.

Definition 3.7. (*Cumulative Distribution Function*)
*The **c**umulative **d**istribution **f**unction (cdf) of a continuous random variable $X \sim F$ is defined by*

$$F(x) := \mathbb{P}(X \le x) = \int_{-\infty}^{x} f(x)\, dx, \ for \ x \in \mathbb{R},$$

where f denotes pdf of distribution F.

Thus, $F(x)$ denotes the probability that the realization of a random variable X is less or equal than the value $x \in \mathbb{R}$ (i.e. $F(x) = \mathbb{P}(X \le x)$). In contrast, the theoretical quantile function is the inverse of the *cdf*.

Definition 3.8. (*Theoretical Quantile Function*)
The theoretical quantile function regarding a random variable $X \sim F$ is defined by

$$q(p) = F^{-1}(x) \coloneqq \inf \{x \in \mathbb{R} \mid F(x) \geq p\} \; for \; 0 \leq p \leq 1.$$

The theoretical quantile $q_p \coloneqq q(p)$ represents a border that splits all possible realization values of random variable X into two classes. If $X \sim F$, then $(100 \cdot p)\%$ of all realizations of X can be expected to be smaller or equal than q_p. Accordingly, in $(100 \cdot (1-p))\%$ of all cases the realizations can be expected to be larger or equal than q_p. With regard to a normally distributed random variable, that is a feature $X \sim N(\mu, \sigma^2)$, it is assumed that the process mean μ corresponds approximately to the nominal value m. For skewed distributions with heavy tails, it is more robust to design the production process such that the process median corresponds to the nominal value. The median of a random variable X, which refers to $q_{0.5}$, ensures that the nominal value is in the center of all realizations. The alignment of the process median with the nominal value is consistent with the normal distribution, where $\mu = q_{0.5}$ holds. A closed form of the median does not exist for the Beta distribution, as provided by definition 3.6. Nevertheless, the median of a Beta distributed random variable X can be approximated by the distribution parameters α and β in the following way (cf. Kerman [2011]).

$$q_{0.5} \approx \frac{\alpha - \frac{1}{3}}{\alpha + \beta - \frac{2}{3}} \; \text{for} \; \alpha, \beta > 0. \tag{3.10}$$

The definition of a capable production process, as provided by Definition 3.2, shall be generalized by the aid of theoretical quantiles.

Definition 3.9. (*Generalized Process Capability*)

A production process is generally denoted as capable, if all monitored features are random variables $X_{\breve{j}} \sim F_{\breve{j}}$ complying with

$$\forall \breve{j} : LSL_{\breve{j}} \overset{!}{\leq} q_{\breve{j}0.00003} \; and \; USL_{\breve{j}} \overset{!}{\geq} q_{\breve{j}0.99997} \; and \; q_{\breve{j}0.5} \approx m_{\breve{j}}. \tag{3.11}$$

Thus, the tolerance intervals of all monitored features have to possess at least 99.994% of the realized values. If $F_{\breve{j}} \equiv N(\mu_{\breve{j}}, \sigma_{\breve{j}}^2)$ holds for all $\breve{j}$, definition 3.9 corresponds to definition 3.2. Provided that only the specification limits $LSL_{\breve{j}}$ and $USL_{\breve{j}}$ of a feature $X_{\breve{j}}$ are given, the Beta distribution is proposed when one of the limits is further away from the nominal value than the other. The methodology on how to determine the parameters $\alpha_{\breve{j}}$ and $\beta_{\breve{j}}$ of the Beta distribution, when the production process of $X_{\breve{j}}$ is barely capable, that is $LSL_{\breve{j}} = q_{\breve{j}0.00003}$ and $USL_{\breve{j}} = q_{\breve{j}0.99997}$, is shown in subsection 3.1.3 by example 3.6. Manufacturing tolerances concerning a feature $X_{\breve{j}} \sim F_{\breve{j}}$ are commonly outlined by a tolerance interval $T_{\breve{j}} = [LSL_{\breve{j}}, USL_{\breve{j}}]$. While the spanning specification limits are frequently given, information about the process capability or the distribution $F_{\breve{j}}$ is rarely available. This gap of knowledge concerning the unknown distribution $F_{\breve{j}}$ is likely bridged by the normal distribution for two reasons. At first, the assumption of the normal distribution often turns out to be adequate because of the CLT. In addition to that, even if skewed features are apparently indicated, the normal distribution is consciously accepted by many engineers. This is because of the methodical advantages of the normal distribution, given by invariance to convolution, "easy to have" parameter estimators, universal acceptation and broad application in engineering tasks. In spite of the methodical advantages

of the normal distribution, the accurate modeling of existent manufacturing tolerances shall be emphasized in this thesis. Consequently, the versatile Beta distribution shall be applied, if skewed data is clearly indicated. Independent of the chosen distribution type, it is suggested to assume barely capable production processes, specifically given by $LSL_{\check{j}} = q_{\check{j}0.00003}$ and $USL_{\check{j}} = q_{\check{j}0.99997}$ for all $\check{j}$. Thus, the monitored feature $X_{\check{j}}$ realizes between $LSL_{\check{j}}$ and $USL_{\check{j}}$ in approximately 99.994% of all cases. Simulating manufacturing tolerances requires continuously accessing and modifying the right parameters in MoBEO. Yet, it needs to be clarified whether or not MoBEO is capable of processing a given manufacturing tolerance information by its parameters.

3.1.2 Processing Manufacturing Tolerances

In theory, manufacturing tolerances refer to hardware features $X_{\check{j}}$ $(\check{j} = 1, \ldots, \check{k})$ that are subject to distributions $F_{\check{j}}$. Still, manufacturing tolerances are usually outlined by a nominal design value $m_{\check{j}}$ and specification limits $LSL_{\check{j}}$ as well as $USL_{\check{j}}$. As outlined in subsection 3.1.1, $F_{\check{j}}$ can be obtained by assuming a bare process capability so that the available values correspond to the theoretical quantiles $m_{\check{j}} = q_{\check{j}0.5}$, $LSL_{\check{j}} = q_{\check{j}0.00003}$ and $USL_{\check{j}} = q_{\check{j}0.99997}$. Simulating manufacturing tolerances of $X_{\check{j}} \sim F_{\check{j}}$ requires accessing and modifying the right parameters in MoBEO. If available manufacturing tolerances are not processable by parameters, it is required to either transform the available information or gather new manufacturing tolerance information that is processable. It is the semi-physical structure of MoBEO that makes a careful processing of available manufacturing tolerance information necessary.

The semi-physical structure of MoBEO in part permits accessing hardware features of the engine environment by parameters. This is, for example, the case for the geometry of individual EAS components, where a monolith diameter or length is processed in a physical way. Given that $X_{\check{j}} \sim F_{\check{j}}$ denotes a processable hardware feature, manufacturing tolerances of $X_{\check{j}}$ can be simulated by setting the corresponding MoBEO parameter on realization values $x_{\check{j}}$ in the forefront of the simulation. Regarding the diameter of an EAS catalyst, for example, manufacturing tolerances could be outlined by $0.15m \pm 1.75mm$. In correspondence to equation (3.8), manufacturing tolerances are simulated by setting the diameter to realization values of $X \sim N(0.00015\,mm, (1.75\,mm/4)^2)$ in the MoBEO model.

MoBEO does not allow the consideration of manufacturing tolerances for all parts at the very lowest hardware level. Nevertheless, although manufacturing tolerances theoretically aim hardware features, a detailed consideration of the tiniest components is not always reasonable in engineering tasks. To give an example, a prospective customer of an expensive watch is certainly more interested in the temporal preciseness than in dimension deviations, chemical impurities or location tolerances of every single component of the clockwork. In almost the same manner this applies to components obstructed in a vehicle, just as sensors or actuators. There, the overall accuracy is usually far more important to engineers than their exact composition. Indeed, tolerance information made available for these components has almost exclusively of functionality inaccuracies. Sensors, for instance, establish the important link between the ECU and physical as well as chemical processes during engine operation. The relationship between the stimulus S of a sensor,

which may be a measured quantity, a condition, or a property, and the according output signal is described by a so called *transfer function* τ that is usually unknown. For simplicity, τ is frequently assumed in a simple parametric form, given by

$$\tau(S) = \Delta + \Gamma \cdot S \text{ where } \Delta, \Gamma \in \mathbb{R}. \tag{3.12}$$

While the intercept Δ refers to the transmitted signal, when stimulus $S = 0$, the slope parameter Γ refers to the influence of $S \neq 0$. In practice, Δ and Γ are unknown and have to be determined by a calibration procedure, where the output signal of a few random samples with known reference source is investigated. Once, Δ and Γ have been estimated, the ECU is able to process the output signal of the sensor by the inverse transfer function. In most cases, the form (3.12) is only a rough approximation of the real transfer function that always consists at least a small portion of non-linearity. Nevertheless, higher calibration costs often deter manufactures from considering more complex transfer functions. Additionally, as pointed out by Fraden [2010], the intercept and slope are not always estimated for every individual sensor so that production tolerances may cause the ECU to 'misinterpret' the received output signals. Hence, the intercept δ and slope γ of the actual obstructed sensors do not necessarily correspond to the estimated intercept and slope parameter m_Δ and m_Γ, deposited in the ECU. The following sensor errors are distinguished in measurement technology (cf. Reif [2012]).

1. The *offset error* combines all error sources of the production and the calibration procedure. If a linear transfer function is used, the offset error denotes a deviation from the deposited intercept m_Δ, causing a systematic bias of the output signal (cf. figure 3.4). Hence, the offset error refers to a continuous random variable $\Delta \sim F_\Delta$, where $m_\Delta = q_{\Delta_{0.5}}$, $LSL_\Delta = q_{\Delta_{0.00003}}$ and $USL_\Delta = q_{\Delta_{0.99997}}$.

2. The *gain error*, which is intended to cover deviations of the resistors within the circuit, is a special case of the linear transfer function and denotes a deviation from the deposited slope parameter m_Γ (cf. figure 3.5). The gain error refers to a continuous random variable $\Gamma \sim F_\Gamma$, where $m_\Gamma = q_{\Gamma_{0.5}}$, $LSL_\Gamma = q_{\Gamma_{0.00003}}$ and $USL_\Gamma = q_{\Delta_{0.99997}}$.

3. Physical peculiarities of the sensor cause the *non-linearity error*. The non-linearity error denotes a conditional random variable $\Lambda|S$, whose distribution depends on the sensor stimulus S. Consequently, $m_{\Lambda|S} = q_{\Lambda|S_{0.5}}$, $LSL_{\Lambda|S} = q_{\Lambda|S_{0.00003}}$ and $USL_{\Lambda|S} = q_{\Lambda|S_{0.99997}}$ correspond to functions.

The simulation of the offset and gain error is easily accomplished, because linear transfer functions, as given by (3.12), are available in MoBEO. Thus, it suffices to alter the respective offset and gain parameter on account of the available tolerance information. That is setting the parameter δ and γ regarding the distributions F_Δ and F_Γ. As illustrated by figure 3.6, the non-linearity error denotes a quantitative factor, which describes a percentage deviation between input and output signal that depends on the sensor stimulus S. It is a characteristic of the non-linearity error that it cannot be corrected by a calibration procedure using a linear transfer function. Simulating non-linearity errors requires considering conditional random variables in terms of manufacturing tolerances.
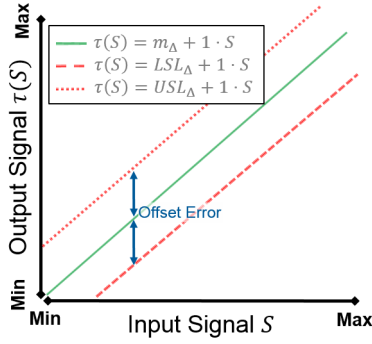
Figure 3.4: Offset Error
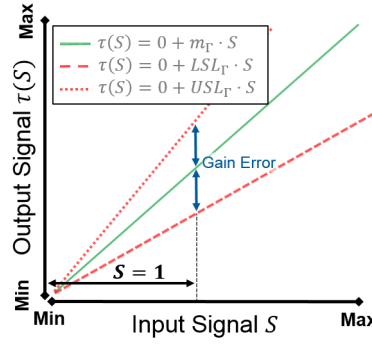
Figure 3.5: Gain Error
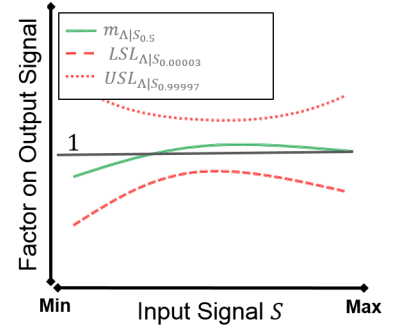
Figure 3.6: Non-Linearity Error

Definition 3.10. (*Independent Random Variables*)

Two random variables $X_i \sim F_i$ and $X_j \sim F_j$ $(i \neq j)$ are called (stochastically) independent, if and only if

$$\mathbb{P}\left(X_i \leq x_i, X_j \leq x_j\right) = F_i(x_i) \cdot F_j(x_j) \ \text{for all} \ (x_i, x_j) \in \mathbb{R}^2. \tag{3.13}$$

Likewise, the random variables $X_1, \ldots, X_k$ are independent, if and only if

$$\mathbb{P}\left(X_1 \leq x_1, \ldots, X_k \leq x_k\right) = \prod_{j=1}^{k} F_j(x_j) \ \text{for all} \ \mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k. \tag{3.14}$$

The joint cdf of k independent random variables $X_1, \ldots, X_k$ is given by

$$F_{(1,\ldots,k)}^{k}(\mathbf{x}) = \prod_{j=1}^{k} F_j(x_j) \ \text{for all} \ \mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k. \tag{3.15}$$

The notation F^k highlights the dimension of the underlying distribution. In probability theory, the chain rule gives the general definition for the joint *cdf* of two random variables $X_i \sim F_i$ and $X_j \sim F_j$ that are not necessarily independent (cf. Bishop [2007]).

Definition 3.11. (*Conditional Random Variables*)

The joint cdf of two not necessarily independent (conditional) random variables X_i and X_j is given by

$$F_{(i,j)}^{2}(\mathbf{x}) = F_{X_i|X_j=x_j}(x_i) \cdot F_j(x_j), \tag{3.16}$$

where

$$F_{X_{\tilde{i}}|X_{\tilde{j}}=x_{\tilde{j}}}(x_{\tilde{i}}) = \mathbb{P}\left(X_{\tilde{i}} \leq x_{\tilde{i}} \mid X_{\tilde{j}} = x_{\tilde{j}}\right) \tag{3.17}$$

denotes the conditional cdf of random variable X_i given $X_j = x_j$ for all $(x_i, x_j) \in \mathbb{R}^2$.

It is also possible that a random variable X_i depends on more than one random variable. Then, one may decompose the joint distribution as follows

$$F_{(i,j)}^{k}(\mathbf{x}) = F_{X_i|\mathbf{X}_{j\neq i}=\mathbf{x}_{j\neq i}}(x_i) \cdot F_{j\neq i}^{k-1}(\mathbf{x}_{j\neq i}), \tag{3.18}$$

where

$$F_{X_i|X_j=x_j}(x_i) = F_{X_i|\mathbf{X}_{j\neq i}=\mathbf{x}_{j\neq i}}(x_j) \tag{3.19}$$

denotes the conditional *cdf* of random variable X_i given $\mathbf{X}_{j\neq i} := (x_j)_{j\neq i}$ for all $\mathbf{x} \in \mathbb{R}^k$. The non-linearity error is simulated in MoBEO by multiplying the output signal $\tau(S)$ with realization values of the conditional random variable $\Lambda|S \sim F_{\Lambda|S}$ that depends on engine operation. Realized values s of S may depend on other simulated manufacturing tolerances so that Λ is not necessarily independent of other $X_{\tilde{j}}$ simulated. Especially, this concerns all random variables which have an influence on the stimulus S. Nevertheless, an independent simulation of the non-linearity error can be ensured by implementing the functions $m_{\Lambda|S}$, $LSL_{\Lambda|S}$ and $USL_{\Lambda|S}$ into MoBEO. This is shown by the following example.

Example 3.1. (*Simulation of the Non-Linearity Error*)

In measurement technology, the NO_x sensor is a probably well known representative in terms of non-linearity errors. The measurement accuracy Λ of a NO_x sensor typically depends on the NO_x concentration S in the exhaust gas. In particular, the accuracy specification of a state-of-the-art NO_x sensor is given by $m_{\Lambda|S} = m_\Lambda = 0\,ppm$ and

$$LSL_{\Lambda|S} = \begin{cases} -10\,ppm & \text{if } S \leq 10\,ppm \\ -10\,\% & \text{if } S > 10\,ppm \end{cases}, USL_{\Lambda|S} = \begin{cases} +10\,ppm & \text{if } S \leq 10\,ppm \\ +10\,\% & \text{if } S > 10\,ppm \end{cases}, \tag{3.20}$$

as illustrated by figure 3.7. According to subsection 3.1.1, the equidistant specification lim-
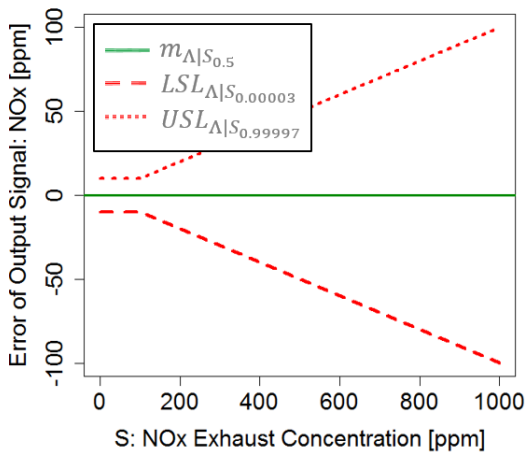
Figure 3.7: NO_x Sensor: Non-Linearity Error
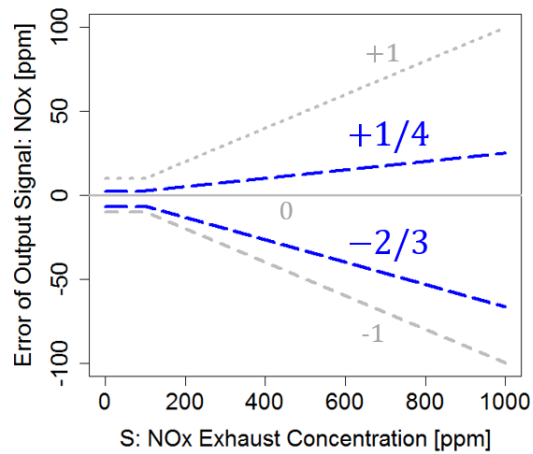
Figure 3.8: NO_x Sensor: Simulation of the Non-Linearity Error

its propose the modeling by a normal distributed random variable, subject to $N(0, \sigma^2(S))$. Consequently, the introduced approach on how to simulate manufacturing tolerances fails. This is because setting the corresponding MoBEO parameter to the realized value of a $N(0, \sigma^2(S))$ distributed random variable is not possible in the forefront of the simulation. This problem can be resolved, however, if the functional relationship of the non-linearity

*error is implemented into MoBEO. In practice this means, that a dummy parameter X
is introduced, which may be chosen within the interval $[-1, +1]$. Setting the dummy pa-
rameter to "$x = 0$" before starting the simulation, means for example that a sensor is
simulated, whose non-linearity error corresponds to $m_\Lambda = 0\,ppm$. In contrast, setting the
parameter to "-1" ("$+1$") means that a sensor is simulated, whose accuracy exactly equals
to $LSL_{\Lambda|S}$ ($USL_{\Lambda|S}$). If a real value between "-1" and "$+1$" is chosen, the sensor ac-
curacy shall correspond to the respective linearly interpolated form between $LSL_{\Lambda|S}$ and
$USL_{\Lambda|S}$. Thus, as illustrated for the dummy values "$-\frac{2}{3}$" and "$+\frac{1}{3}$" by figure 3.5, the
assumption has to be made that the functional error form of an arbitrary sensor reflects
the specification limits given. To sum up, the non-linearity error of the NO_x sensor can be
modeled by the random variable $X \sim N\left(0, \frac{1}{4}\right)$. It can be independently simulated by setting
the introduced dummy parameter to the realization values of X, once the functional error
forms are implemented in MoBEO.*

In principle, the simulation of the offset and the gain error would require similar mea-
sures, as presented for the non-linearity error. Due to the already implemented transfer
function (3.12) this is, however, not necessary, because the offset and gain error directly
correspond to the dummy variable introduced. Among others the non-linearity error of
a sensor may depend on manufacturing tolerances of actuators, used to convert electri-
cal signals into mechanical labor. Comparable with sensors, manufacturing tolerances
concerning actuators are usually evaluated on how accurately they operate, but not in
regard of any hardware features. In fact, the simulation of manufacturing tolerances of
an actuator corresponds to the simulation of a non-linearity error.
Although manufacturing tolerances theoretically refer to hardware features, engineers may
rather prefer assessing the overall functionality of a monitored component. Just as for
sensors and actuators, this simplification accommodates the simulation of manufacturing
tolerances in MoBEO, where not every hardware feature is considerable. If available man-
ufacturing tolerance information is not processable by parameters, there exists at least
the chance to transform the available information in an appropriate manner, as shown by
the subsequent example.

Example 3.2. (*Mathematical Transformation of Manufacturing Tolerances*)

*The crankshaft position sensor is an important device in the powertrain that supplies the
ECU with information about the crankshaft's current rotational speed. An economical fuel
injection as well as an accurate boost pressure and EGR control are realized with this
information. The following offset errors are available for a certain crankshaft position
sensor. The equidistant specification limits propose the assumption of the normal distri-
bution, as described by (3.8). Although it is not possible to individually process these offset
errors by MoBEO, the information of table 3.1 can be summed up by convolution, if the
errors are assumed as stochastically independent. The root sum of squares approach, as
described by Scholz [1995], gives an overall offset error that can be processed.*

Beside mathematical transformation techniques also heuristic transformations can be
applied. Many physical processes are modeled by empirical auxiliary models in MoBEO.
Their complexities reach from simple constant functions to multivariate mathematical
models, whose parameters may take special physical roles.

Tolerance Cause	Tolerance	Classification	Responsibility
Evaluation Circuit & Gate Array in Control Unit	$\pm 0.15°CA$	Offset	Manufacturer
Sensing Element	$\pm 0.50°CA$	Offset	Manufacturer
Installation Accuracy (Shift from TDC)	$\pm 0.25°CA$	Offset	Customer
Tone Wheel Installation Accuracy (Shift from TDC)	$\pm 0.20°CA$	Offset	Customer
Tooth Tolerance (Spacing Error, Flanks)	$\pm 0.10°CA$	Offset	Customer
Engine (Crank Tolerance, etc.)	$\pm 0.70°CA$	Offset	Customer
Overall Offset Error (Root Sum of Squares)	$\pm 0.875°CA$	Offset	Manufacturer/Customer

Table 3.1: Production Tolerances: Crankshaft Position Sensor

Example 3.3. (*Heuristic Transformation of Manufacturing Tolerances*)

The Arrhenius' equation approach (cf. Heck et al. [2009]), for instance, is used to model chemical reactions occurring inside an EAS catalyst. This equation describes the rate constant K of a reaction to be inversely related to the exponential of the activation energy E. In detail, the Arrhenius' equation is given by

$$K = k_0 \cdot \exp\left(-\frac{E}{R \cdot T}\right), \tag{3.21}$$

where T is the absolute temperature, R the universal gas constant and k_0 the preexponential constant. The parameters of the Arrhenius' equation are set in a way that measurement data, made with a real catalyst, is well interpolated. As a result, manufacturing tolerances with regard to features like precious metal coating or washcoat loading are hidden in the parameters of Arrhenius' equation. Simulating manufacturing tolerances of such hidden features is realized by adequately transmitting the available tolerance information to the parameters of the empirical auxiliary model. On this account, let $X \sim F$ denote the feature precious metal coating, whose manufacturing tolerances are given. Assuming that $m \neq 0$, it is proposed to alter the preexponential constant k_0 of the respective Arrhenius' equation within the transformed tolerance interval $[k_0 \cdot LSL/m, k_0 \cdot USL/m]$ in 99.9994% of all cases. This is because the preexponential constant is assumed to be proportional to the catalytically active surface of the catalyst.

In practice, one Arrhenius' equation is used for every chemical reaction modeled. Thus, with regard to a **S**elective **C**atalytic **R**eduction (SCR) catalyst for instance, the HC-oxidation and NO-oxidation are modeled by separate Arrhenius' equations. If manufacturing tolerances of the feature *precious metal coating* are considered by the respective preexponential constants, it would be reasonable that these depend on each other. This issue will be reconsidered in example 3.9 in the next subsection.

Just as for precious metal coating, hardware features of the **C**harge **A**ir **C**ooler (CAC), like the overall geometry or the quality of the materials used to build the core and header plates, are not considerable by MoBEO parameters. These hardware features are instead hidden by an empirical auxiliary model that determines the CAC hot effectiveness η_{CAC} at different intake air mass flows. The CAC hot effectiveness is defined by

$$\eta_{CAC} = \frac{T_{IN} - T_{OUT}}{T_{IN} - T_{Coolant}}, \tag{3.22}$$

where T_{IN} (T_{OUT}) denotes the absolute air temperature before (after) the CAC, and $T_{Coolant}$ refers to the absolute temperature of the coolant. Regarding η_{CAC}, neither a direct processing of the available manufacturing information, aiming the overall geometry as well as quality, nor a mathematical or heuristic transformation thereof is applicable. Consequently, manufacturing tolerances of the CAC can only be considered by MoBEO, if manufacturing tolerances are re-assessed in terms of η_{CAC}. This means, a random sample of charge air coolers has to be investigated, and the distribution $F_{\eta_{CAC}}$ of the random variable η_{CAC} needs to be estimated at different intake air mass flows. Once, the specification limits and the nominal values of the η_{CAC} are estimated for a set of support points, a linear interpolation of the acquired tolerance information can be applied (cf. figure 3.9). The simulation of manufacturing tolerances concerning the CAC is realized,
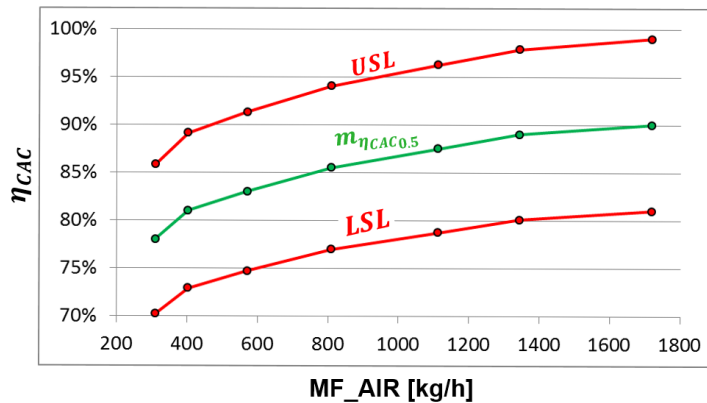
Figure 3.9: Functionality Tolerance of the CAC

just as for the non-linearity error of a sensor. Thus, the functional forms obtained need to be implemented in MoBEO such that they can be controlled by a dummy parameter $X \sim N(0, \frac{1}{4})$.

To sum up, the reason is twofold that manufacturing tolerances are not always considered concerning hardware features $X_1, \ldots, X_{\tilde{k}}$. On the one hand, MoBEO uses empirical auxiliary models, like the Arrhenius' equation, so that processing tolerance information aiming hardware features is not always directly possible. On the other hand, manufacturing tolerances are not always outlined in terms of hardware features. This is, for example, the case for more complex devices of the powertrain, like sensors or actuators, where the functionality but not hardware features are assessed. In the end, it does not really count whether or not hardware features are aimed. It is the right interpretation and processing of the available tolerance information that must be overcome. The simulation of manufacturing tolerances is basically realized by modifying parameters in MoBEO. For that reason, the continuous random variables $X_1 \sim F_1, \ldots, X_{\tilde{k}} \sim F_{\tilde{k}}$ should from now on refer to parameters, which are modified in MoBEO. Hence, manufacturing tolerances are considered by MoBEO parameters, but not necessarily by hardware features. As a consequence of this subsection, $X_1, \ldots, X_{\tilde{k}}$ cannot be generally assumed as independent random variables. The original distributions concerning hardware features, if available, need to be either adequately transformed or re-estimated out of adequate measurement data. On this account, the next subsection will present procedures on how to estimate

unknown univariate and multivariate distributions.

3.1.3   Estimation Procedures and the Nominal Vehicle

Simulating manufacturing tolerances by MoBEO requires a careful interpretation and processing of a possibly available information. Basically, manufacturing tolerances, which are considered by a continuous random variable $X_{\breve{j}} \sim F_{\breve{j}}$, are simulated by setting the according MoBEO parameter to realized values $x_{\breve{j}}$ for $\breve{j} = 1, \ldots, \breve{k}$. As outlined in subsection 3.1.1 and 3.1.2 a possibly unknown distribution $F_{\breve{j}}$ can be obtained by transformation procedures, taking into account available specification limits for example. This subsection will focus on the case where the required distribution $F_{\breve{j}}$ cannot be obtained by the information available. Then, $F_{\breve{j}}$ may be estimated by *parametric estimation methods*, by *nonparametric estimation methods* or by a mixture of both approaches (cf. Bishop [1995]). With regard to subsection 3.1.1, it seems to be favorably to use parametric estimation methods. On this account, it is proposed to generally use parametric estimation methods for manufacturing tolerances. Although, most of the considered random variables $X_1, \ldots, X_{\breve{k}}$ are independent in practice, the semi-physical structure of MoBEO makes it necessary to additionally consider multivariate distributions. At the end of this subsection, the model of the so called *nominal vehicle*, is introduced.

Let the observations $\mathbf{x}_{\breve{j}} = \left\{ x_{1\breve{j}}, \ldots, x_{\breve{N}\breve{j}} \right\}$ denote a random sample of a random variable $X_{\breve{j}}$. The agreement on a certain parametric distribution type $F_{\breve{j}}(\boldsymbol{\theta}_{\breve{j}})$ and the subsequent estimation of the parameter vector $\boldsymbol{\theta}_{\breve{j}}$ on basis $\mathbf{x}_{\breve{j}}$ is called *parametric estimation approach*. In accordance with subsection 3.1.1, the predetermined distribution types shall be either the normal distribution with parameter vector $\boldsymbol{\theta}_{\breve{j}} = \left(\mu_{\breve{j}}, \sigma_{\breve{j}}^2 \right)$ or the versatile Beta distribution with parameter vector $\boldsymbol{\theta}_{\breve{j}} = \left(\alpha_{\breve{j}}, \beta_{\breve{j}} \right)$. A simple histogram, a boxplot or statistical tests (e.g. *Shapiro-Wilk test*) of the monitored random sample $\mathbf{x}_{\breve{j}}$ can be used to decide on one of the two distribution types. The most popular parametric methods estimating $\boldsymbol{\theta}_{\breve{j}}$ are the *method of moments*, the *maximum likelihood method* and *least-squares estimation method* (cf. Casella and Berger [2006]). In case of the normal distribution, the respective estimators are equivalent for all of these parametric methods. The unknown parameters $\mu_{\breve{j}}$ and $\sigma_{\breve{j}}^2$ are estimated as follows,

$$\mu_{\breve{j}} \approx \bar{x}_{\breve{j}} = \frac{1}{\breve{N}} \cdot \sum_{\breve{n}=1}^{\breve{N}} x_{\breve{n}\breve{j}} \text{ and } \sigma_{\breve{j}}^2 \approx s_{\breve{j}}^2 = \frac{1}{\breve{N}} \cdot \sum_{\breve{n}=1}^{\breve{N}} \left(x_{\breve{n}\breve{j}} - \bar{x}_{\breve{j}} \right)^2 .^2 \qquad (3.23)$$

As shown in subsection 3.1.1, theoretical quantiles are essential to define production capability, if $F_{\breve{j}}$ is not a normal distribution. Assuming a bare process capability, the nominal value $m_{\breve{j}}$ and specification limits $LSL_{\breve{j}}$ as well as $USL_{\breve{j}}$ correspond to the theoretical quantiles $q_{\breve{j}0.5}$ and $q_{\breve{j}0.00003}$ as well as $q_{\breve{j}0.99997}$. If $F_{\breve{j}}$ is unknown, these theoretical quantiles need to be estimated by *empirical quantiles*. Therefore, let $x_{(1\breve{j})} \leq \cdots \leq x_{(\breve{N}\breve{j})}$ denote the realization of the *order statistic* $X_{(1\breve{j})} \leq \cdots \leq X_{(\breve{N}\breve{j})}$.

Definition 3.12. (*Empirical Distribution Function*)
*The **e**mpirical **d**istribution **f**unction (edf) of a random variable X is determined by a*

---

^{2}Often estimated by the unbiased estimator $s_{\breve{j}}^2 = \frac{1}{\breve{N}-1} \cdot \sum_{\breve{n}=1}^{\breve{N}} \left(x_{\breve{n}\breve{j}} - \bar{x}_{\breve{j}} \right)^2$.

corresponding random sample $\mathbf{x} = \{x_1, \ldots, x_N\}$ *as follows*

$$F_N(x) := \begin{cases} 0 & x < x_{(1)} \\ \frac{n}{N+1} & x_{(n)} \leq x < x_{(n+1)} \quad \text{for } n = 1, \ldots, N-1 \\ 1 & x_{(N)} \leq x \end{cases} \tag{3.24}$$

It can be shown that (cf. Rohatgi [1976])

$$F_N(x) \xrightarrow{N \to \infty} F(x) \ \forall x \in \mathbb{R}. \tag{3.25}$$

The empirical quantile function can be seen as the inverse of the *edf*.

Definition 3.13. (*Empirical Quantile Function*)
The p-th theoretical quantile q_p of a distribution F satisfies $F(q_p) = p$ and respectively $q_p = F^{-1}(p)$ for any probability $0 < p < 1$. The p-th empirical quantile x_p of a random sample $\mathbf{x} = \{x_1, \ldots, x_N\}$, estimating q_p, can be determined by the ordered random sample $x_{(1)} \leq \cdots \leq x_{(N)}$ as follows (cf. Hyndman and Fan [1996])

$$x_p(\mathbf{x}) = x_p := (1 - g) \cdot x_{(\lfloor (N-1) \cdot p \rfloor + 1)} + g \cdot x_{(\lfloor (N-1) \cdot p \rfloor + 2)}, \tag{3.26}$$

where $g = (N-1) \cdot p - \lfloor (N-1) \cdot p \rfloor$.

Example 3.4. (*Estimation of the Normal Distribution on Basis of Measurement Data*)

The isentropic compressor efficiency is the ratio between the minimum work necessary and the actual work required by a compressor unit to increase the pressure of the intake air mass flow to a certain level. Due to manufacturing tolerances, identically constructed compressor units may exhibit different isentropic compressor efficiencies at approximately identical air mass flows. On this account, $\check{N} = 15$ equally constructed turbo chargers have been investigated in terms of their isentropic compressor efficiencies. As shown by figure 3.10, each compressor has been investigated at four $(m = 1, \ldots, 4)$ different rotational speed values, each supported by eight $(\check{n}_m = 1, \ldots, 8)$ different corrected intake air mass flows. Hence, 32 different random samples $\{x_{1m,\check{n}_m}, \ldots, x_{15m,\check{n}_m}\}$ are available. If these are standardized by the empirical medians

$$x_{\check{n}m,\check{n}_m}^{std} = \frac{x_{\check{n}m,\check{n}_m}}{x_{0.5}\left(\left\{x_{1m,\check{n}_m}, \ldots, x_{15m,\check{n}_m}\right\}\right)} \tag{3.27}$$

for all $\check{n} = 1, \ldots, 15$, $m = 1, \ldots, 4$, $\check{n}_m = 1, \ldots, 8$, it is possible to assess the measurement data independently of the intake air mass flow and the rotational speed. Consequently, a random sample of $15 \cdot 4 \cdot 8 = 480$ observations is obtained, whose unknown distribution F shall be estimated. The corresponding histogram, provided by figure 3.11, approximately shows the shape of a normal distribution's density. Assuming that $F \equiv N(\mu, \sigma^2)$, parameters μ and σ^2 have to be determined by equation (3.23). Figure 3.11 also contains the density curve of the resulting normal distribution with parameters $\mu \approx 0.9989$ and $\sigma^2 \approx 0.0215^2$ and the empirical quantiles $x_p(\{x_1^{std}, \ldots, x_{480}^{std}\})$ that estimate the nominal value m and the specification limits LSL and USL for $p \in \{0.5, 0.00003, 0.99997\}$. It is
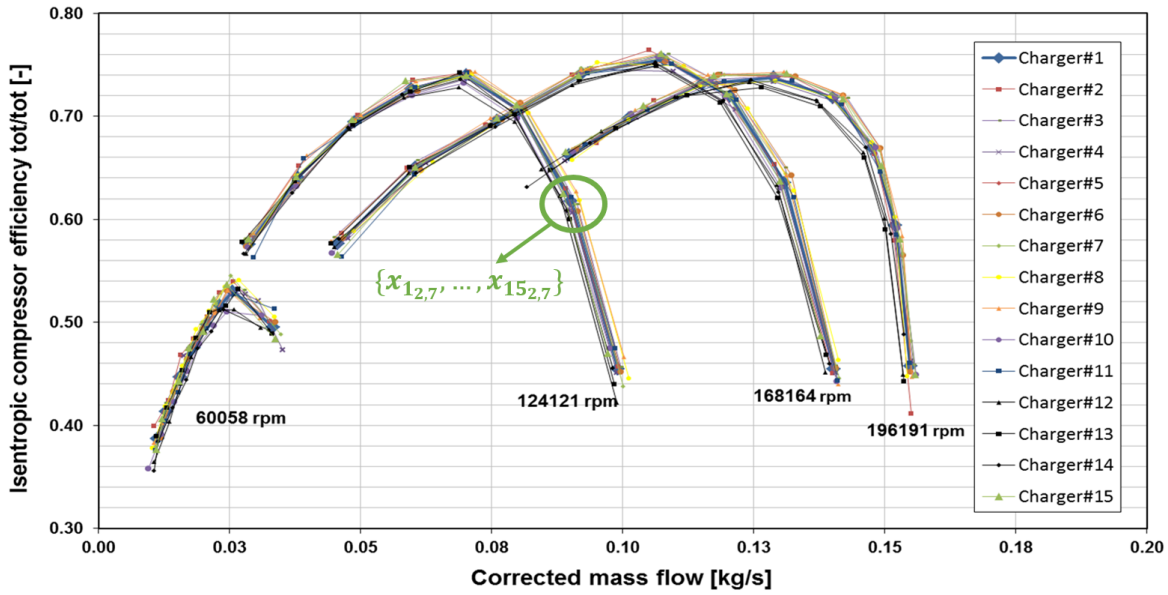
Figure 3.10: Production Tolerance: Compressor Efficiency

clearly visible that the estimated normal distribution not adequate, because too many observations have been made in regions where the density is almost zero. In addition to that, one observation is even larger than the USL, although the expected number of observations being larger than the USL is 0.0144 for 480 observations.

The defective estimation in example 3.4 can be eliminated, if the estimated standard deviation is corrected by the sample size $\check{N}$. This amendment is accomplished by a *one-sided* $(1-\alpha)\%$ *confidence interval* that contains the true variance with a probability $1-\alpha$ for $0 < \alpha \le 0.2$. The one-sided confidence interval, providing an upper bound for the unknown variance, stands in the main focus. For a random sample $\mathbf{x}_{\check{j}} = \left\{ x_{1\check{j}}, \ldots, x_{\check{N}\check{j}} \right\}$ that is the interval

$$\left[0, \frac{\left(\check{N} - 1 \right) \cdot s_{\check{j}}^2}{\chi_{\check{N}-1;\alpha}^2} \right], \tag{3.28}$$

where $\chi_{N-1;\alpha}^2$ is the theoretical quantile q_α of the chi-squared distribution with $(N-1)$ **d***egrees of* **f***reedom* (*df*s). It holds that the probability

$$\mathbb{P} \left(\sigma^2 \in \left[0, \frac{\left(\check{N} - 1 \right) \cdot s_{\check{j}}^2}{\chi_{\check{N}-1;\alpha}^2} \right] \right) = 1 - \alpha. \tag{3.29}$$

Example 3.5. *(Continuation of Example 3.4.)*

Assuming $\alpha = 0.05$ *and* $\check{N} = 15$, *the confidence interval of the variance in regard of the standardized isentropic compressor efficiency measurements is given as* $[0, 0.000986]$. *Figure 3.12 shows the resulting normal distribution, if the estimated standard deviation is replaced by the square root of the upper bound of the one-sided 95% confidence interval. Given that the true variance is within the estimated confidence interval, the assumed process capability is not affected. Thus,* $\mathbb{P} \left(x^{std} \in T \,|\, \sigma^2 \in [0, 0.000986] \right) = 0.99994$. *The*
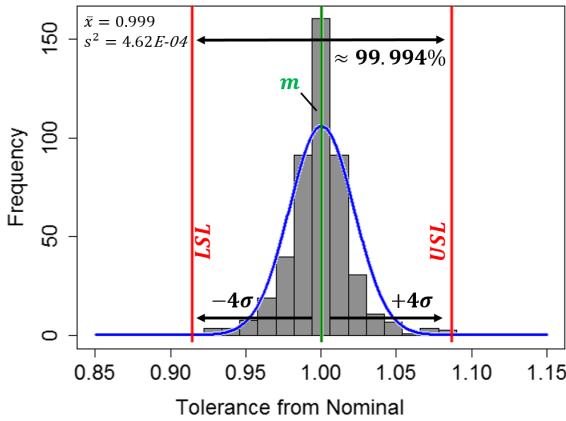
Figure 3.11: Parametric Estimation Approach: Normal Distribution
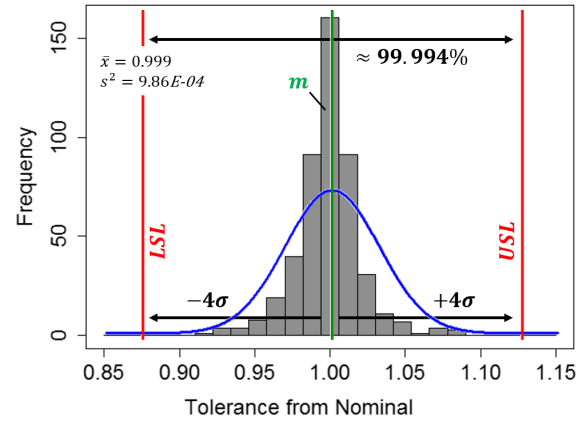
Figure 3.12: Parametric Estimation Approach: Normal Distribution Punished by Sample Size

improvement compared to the normal distribution of figure 3.12 is clearly visible at the tails of the amended normal distribution.

Given that a monitored random sample $\mathbf{x}$ of a random variable X cannot be modeled with the normal distribution, the Beta distribution is assumed for the parametric estimation approach. Abourizk et al. [1994] critically discuss differences between the method of moments, the maximum likelihood method and least-squares estimation method in terms of the general Beta distribution with domain $[a, b] \subset \mathbb{R}$. Its *cdf* is denoted by $F_{X \in [a,b]}(x \mid \alpha, \beta)$. The authors conclude their paper with a favor for the least-squares estimation procedure, because unlike the maximum likelihood method, the least-squares method does not require any knowledge about the unknown domain limits a and b. Moreover, in their paper the least-squares method presented consistently better results than both other methods. Therefore, the least-squares method is presented in this subsection. It uses the non-linear regression model, taking into account the *edf* of $\mathbf{x}$, to estimate the unknown parameter vector $\boldsymbol{\theta} = (\alpha, \beta)$. That is

$$F_{X \in [a,b]}(x_{(\check{n})} \mid \alpha, \beta) = F_{\check{N}}\left(x_{(\check{n})}\right) + \epsilon_{\check{n}} = \frac{\check{n}}{\check{N} + 1} + \epsilon_{\check{n}} \text{ for } \check{n} = 1, \dots, \check{N}, \qquad (3.30)$$

where $x_{(\check{n})}$ denotes the $\check{n}$-th realization of the order statistic $X_{(1)} \leq \cdots \leq X_{(\check{N})}$ and $\epsilon_{\check{n}}$ the model error with expected value $\mathbb{E}(\epsilon_{\check{n}}) = 0$ for all $\check{n}$. Abourizk et al. [1994] suggest to estimate the unknown parameter $\boldsymbol{\theta} = (\alpha, \beta)$ and the domain limits a, b by minimizing the constrained nonlinear optimization problem

$$\min_{\alpha, \beta, a, b} \quad \sum_{\check{n}=1}^{\check{N}} \left(F_{X \in [0,1]}\left(x_{(\check{n})}^{\text{std}} \mid \alpha, \beta\right) - \frac{\check{n}}{\check{N} + 1} \right)^2 \qquad (3.31)$$

$$\text{subject to} \quad \alpha, \beta > 0 \text{ and } a < X_{(1)}, b > X_{(\check{N})},$$

where

$$x_{(\check{n})}^{\text{std}} = \frac{x_{(\check{n})} - a}{b - a} \tag{3.32}$$

are the empirical quantiles, standardized to $[0,1]$. $F_{X \in [0,1]}(x)$ equals the *cdf* of the Beta distribution with domain $[0,1]$, as provided by definition 3.6. Hence, the target function of the optimization problem (3.31) is designed such that the error sum of squares between the theoretical and empirical distribution function is minimized. Nevertheless, it has been observed that the target function of this optimization problem has a disadvantage, because it contains probabilities but not quantiles. As a result, the approach of Abourizk et al. [1994] tends to inaccurate estimation results, especially in regions where the *pdf* is approximately zero. This is problematic for manufacturing tolerances, where improbable specification limits need to be estimated. That problem was resolved, however, by our own approach defining the target function by quantiles. The problem was resolved, however, by a new approach specifically developed during this thesis. There, the target function is simply defined by quantiles so that the optimization problem

$$\min_{\alpha, \beta, a, b} \qquad \sum_{\check{n}=1}^{\check{N}} \left(q_{\frac{\check{n}}{\check{N}+1}} \cdot (b - a) + a - x_{\frac{\check{n}}{\check{N}+1}} \right)^2 \tag{3.33}$$

$$\text{subject to} \quad \alpha, \beta > 0 \text{ and } a < X_{(1)}, b > X_{(\check{N})}$$

minimizes the error sum of squares between the theoretical quantiles

$$q_{\frac{\check{n}}{\check{N}+1}} = F_{X \in [0,1]}^{-1} \left(\frac{\check{n}}{\check{N}+1} | \alpha, \beta \right), \tag{3.34}$$

which are transformed to the domain $[a, b]$, and the standardized empirical quantiles $x_{\frac{\check{n}}{\check{N}+1}}$ for $\check{n} = 1, \dots, \check{N}$. The following example emphasizes the difference between both optimization problems and clarifies the open topic of subsection 3.1.1. That is the estimation of parameters $\alpha_{\check{j}}$ and $\beta_{\check{j}}$, whenever the production process is barely capable (i.e. $LSL_{\check{j}} = q_{\check{j}0.00003}$ and $USL_{\check{j}} = q_{\check{j}0.99997}$) according to a continuous random variable $X_{\check{j}}$.

Example 3.6. *(Estimation of the Beta Distribution on Basis of Specification Limits)*

Manufacturing tolerances concerning the hardware feature precious metal coating can be simulated with MoBEO by the preexponential constant of the respective Arrhenius' equation. Deviations in precious metal coating concern the amount of Platinum (Pt), which supports

$$NO \rightarrow NO_2, CO \rightarrow CO_2, HC \rightarrow CO_2 + H_2O \text{ and } NH_3 \rightarrow N_2 + H_2O \text{ (only ASC)}[3].$$
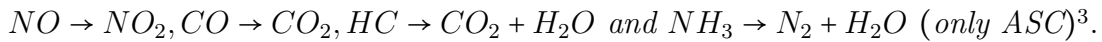
and the amount of Palladium (Pd) that is frequently added for cost as well as thermal stabilization reasons and to support reactions like

$$CO \rightarrow CO_2 \text{ and } HC \rightarrow CO_2 + H_2O.$$

---

[3]cf. section A.3

The ratio between the precious metal coating applied on an arbitrary produced catalyst and the nominal amount of precious metal coating is outlined by the specifications limits $LSL = 0.94$ and $USL = 1.01$. Given that the nominal value corresponds to $m = 1$, the different distances to the specification limits suggest using the Beta distribution. Assuming bare production capability, the theoretical quantiles of the Beta distribution can be estimated by $q_{0.5} \approx m$, $q_{0.00003} \approx LSL$ and $q_{0.99997} \approx USL$. Optimization problems (3.31) and (3.33) are solved by the constrained quasi-Newton method of Byrd et al. [1995] for $x_{(1)} = LSL$, $x_{(2)} = m$ and $x_{(3)} = USL$. The resulting Beta distributions and its theoretical quantiles $q_{0.5}$, $q_{0.00003}$ and $q_{0.99997}$ are provided by figure 3.13. There, it is clearly visible that the optimization problem, used by Abourizk et al. [1994], shows a bad fit for the lower specification limit. By applying the amended target function a Beta distribution is found instead, whose theoretical quantiles $q_{0.5}$, $q_{0.00003}$ and $q_{0.99997}$ almost completely correspond to the preset specification values (cf. figure 3.14).
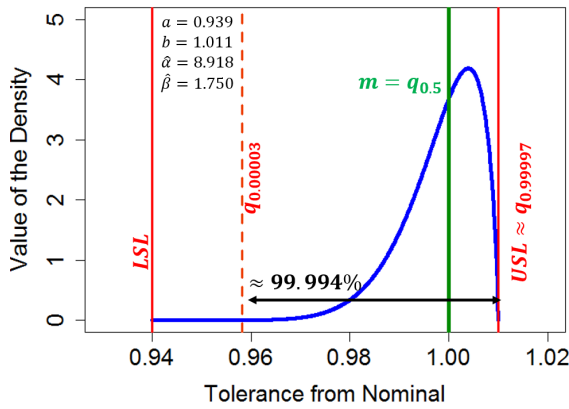
Figure 3.13: Parametric Estimation Approach: Beta Distribution Estimated by Probabilities
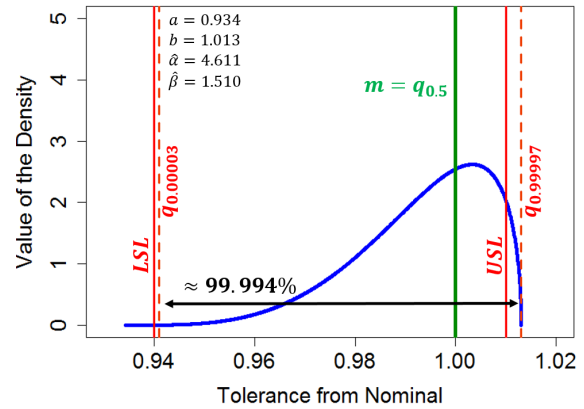
Figure 3.14: Parametric Estimation Approach: Beta Distribution Estimated by Quantiles

For continuous random variables $X_{\check{j}}$, where no tolerance specification limits are available, the nominal value $m_{\check{j}}$ and the specification limits $LSL_{\check{j}}$ and $USL_{\check{j}}$ need to be estimated on basis of a random sample $\mathbf{x}_{\check{j}}$. Then, empirical quantiles $x_{\check{j}\frac{n}{N+1}}$ need to be determined for $\check{n} = 1, \ldots, \check{N}$ so that the amended optimization problem (3.33) can be used to estimate the parameter vector $\boldsymbol{\theta} = (\alpha, \beta)$. Still, attention has to be paid at the tails of the unknown distribution $F_{\check{j}}$, where empirical quantiles are only poor estimators of the theoretical quantiles.

Example 3.7. (*Estimation of the Beta Distribution on Basis of Measurement Data*)

*The **E**xhaust **G**as **R**ecirculation (EGR) technology is widely used in heavy duty diesel applications to reduce NO_x emissions. A high pressure loop system is obstructed in the considered test powertrain, where a portion of the exhaust gas is taken from upstream of the high pressure turbine and is recirculated into the intake air path. The EGR flow is among*

others ensured by the waste gate valve, which maintains a positive pressure difference between the exhaust and the intake manifold. Exhaust in the intake manifold leads to less oxygen and so to lower combustion temperatures, which counteracts the formation of high NO_x concentrations.

Flow measurement data of four different valve positions $m = 1, \ldots, 4$ is available for $\check{N} = 15$ identically designed EGR valves. The MoBEO parameter port flow coefficient $\grave{\alpha}$ has to be altered in order to consider manufacturing tolerances of the EGR valve. The port flow coefficient describes the fraction of any measured and one reference mass flow rate. In this example, the reference air mass flow rate refers to a totally open EGR valve. Formally, $\grave{\alpha}$ is given as

$$\grave{\alpha}\left(\check{n}, m\right) := \frac{\dot{m}_{meas}\left(\check{n}, m\right)}{\dot{m}_{ref}(m)} \text{ for } \check{n} = 1, \ldots, \check{N} \text{ and } m = 1, \ldots, 4. \tag{3.35}$$

The measured port flow coefficients of the investigated EGR valves are provided by table 3.2. As in example 3.4, the measurement data can be analyzed independently of the

Position in [%]	Valve 1	Valve 2	Valve 3	Valve 4	Valve 5	Valve 6	Valve 7	Valve 8	Valve 9	Valve 10	Valve 11	Valve 12	Valve 13	Valve 14	Valve 15
$\beta_1 = 20$	0.104	0.107	0.104	0.096	0.104	0.097	0.094	0.104	0.107	0.098	0.093	0.096	0.104	0.104	0.099
$\beta_2 = 30$	0.164	0.163	0.159	0.158	0.161	0.158	0.149	0.161	0.161	0.164	0.148	0.149	0.158	0.164	0.161
$\beta_3 = 50$	0.270	0.268	0.269	0.258	0.272	0.258	0.262	0.273	0.265	0.265	0.257	0.256	0.261	0.267	0.278
$\beta_4 = 100$	0.454	0.453	0.456	0.446	0.459	0.451	0.459	0.458	0.451	0.457	0.452	0.452	0.448	0.446	0.467

Table 3.2: Measurement Data of Identical Designed EGR Valves

different valve positions, if the data is standardized by the empirical medians as follows.

$$\grave{\alpha}_{std}\left(\check{n}, m\right) = \frac{\grave{\alpha}\left(\check{n}, m\right)}{x_{0.5}\left(\left\{\grave{\alpha}\left(1, m\right), \ldots, \grave{\alpha}\left(\check{N}, m\right)\right\}\right)} \text{ for } \check{n} = 1, \ldots, \check{N} \text{ and } m = 1, \ldots, 4. \tag{3.36}$$

The standardized data is summarized by the histogram, given in figure 3.15. It is assumed that the unknown distribution $F \equiv Beta\left(\alpha, \beta\right)$, because the standardized data is significantly skewed. Optimization problem 3.33 should be applied to estimate the unknown parameter vector $\boldsymbol{\theta} = \left(\alpha, \beta\right)$. Therefore, the necessary theoretical quantiles are estimated by the empirical quantiles, given in (3.26).

Still, too small sample sizes $\check{N}$ might be problematic. Similarly to the normal distribution, punishing the sample size $\check{N}$ is also possible for the Beta distribution. Let $x_{\frac{n}{N+1}}$ denote the theoretical quantiles of a random sample $\mathbf{x}$ of the random variable X that are obtained by the nonparametric estimation approach and used in (3.34). The spread between the specification limits can be increased, if the distance between all empirical quantiles and the median, obtained by (3.26), is increased by the upper bound of the confidence interval in (3.28), that is

$$x_{\frac{\check{n}}{\check{N}+1}}^{\text{corr}} = \left(x_{\frac{\check{n}}{\check{N}+1}} - x_{0.5}\right) \cdot \sqrt{\frac{\check{N} - 1}{\chi^2_{\check{N}-1;\alpha}}} + x_{0.5}. \tag{3.37}$$
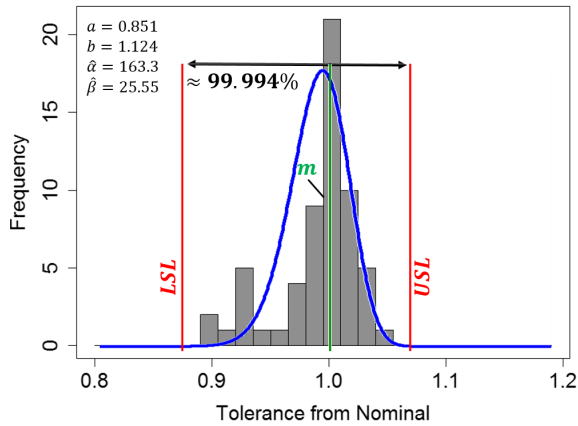
Figure 3.15: Parametric Estimation Approach: Advantage of Robust Quantiles
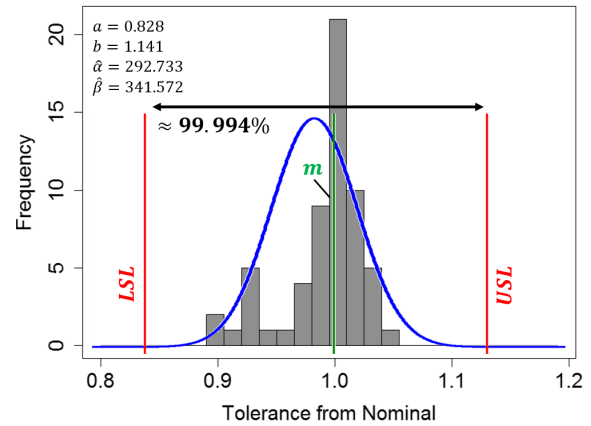
Figure 3.16: Parametric Estimation Approach: Punished by Sample Size

This measure corresponds to an enlargement of the standard deviation as presented in example 3.4. The Beta distribution, which is obtained by applying (3.37) in optimization procedure (3.33), is illustrated in figure 3.16. Dependent on possibly available foreknowledge, it could be sometimes useful to fix the domain $[a, b]$ in the optimization procedure. As outlined in subsection 3.1.2, the continuous random variables $X_1 \sim F_1, \ldots, X_{\check{k}} \sim F_{\check{k}}$, referring to MoBEO parameters, are not necessarily independent. The pairwise assessment of the stochastic relationship between random variables is a common approach in statistics. A measure of this relationship is given by the *covariance*.

Definition 3.14. (*Covariance*)

The covariance of two continuous random variables $X_i \sim F_i$ and $X_j \sim F_j$ is given by

$$COV(X_i, X_j) = \sigma_{ij} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x_i - \mathbb{E}(X_i))(x_j - \mathbb{E}(X_j)) \, f^2_{(i,j)}(x_i, x_j) \, dx_i \, dx_j. \quad (3.38)$$

The sign of the covariance reflects whether the relationship of the random variables is positive or negative. If X_i is independent of X_j, $\sigma_{ij} = 0$ follows. The covariance between two random variables X_i and X_j can be estimated by the *empirical covariance*.

Definition 3.15. (*Empirical Covariance*)

The empirical covariance of two random samples $\mathbf{x}_i = \{x_{1i}, \ldots, x_{Ni}\}$ and $\mathbf{x}_j = \{x_{1j}, \ldots, x_{Nj}\}$ is given by

$$s_{ij} = \frac{1}{N-1} \cdot \sum_{n=1}^{N} (x_{ni} - \bar{x}_i)(x_{nj} - \bar{x}_j). \quad (3.39)$$

For $i = j$ the empirical covariance equals to the unbiased empirical variance (i.e. $s_{jj} = s_j^2$), as outlined in (3.23. A standardized measure is given by *Pearson's correlation coefficient*, which transmits the strength of the relationship between two random samples to the interval $[-1, +1]$.

Definition 3.16. (*Pearson's Correlation Coefficient*)

The Pearson's correlation coefficient r_{ij} of two random samples $\mathbf{x}_i \in \mathbb{R}^N$ and $\mathbf{x}_j \in \mathbb{R}^N$ is given by

$$r_{ij} = \frac{s_{ij}}{s_i \cdot s_j} \;\; with \;\; -1 \le r_{ij} \le +1. \tag{3.40}$$

Due to Fahrmeir et al. [2010], the relationship between two random samples $\mathbf{x}_i$ and $\mathbf{x}_j$ is classified as follows.

Weak correlation: $|r_{ij}| < 0.5$

Medium correlation: $0.5 \le |r_{ij}| < 0.8$

Strong correlation: $0.8 \le |r_{ij}| < 1.0$

Given that two or more correlated random samples represent different normal distributions, the application of a multivariate normal distribution is proposed. Thus, the considered features $X_1, \ldots, X_{\check{k}}$ can be embraced to the random vector $\mathbf{X} = (X_1, \ldots, X_{\check{k}})$ that is subject to a multivariate normal distribution with k dimensions (i.e. $\mathbf{X} \sim N^{\check{k}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^2)$)).

Definition 3.17. (*Density of the $\check{k}$-dimensional Normal Distribution*)

The density function of the multivariate normal distribution is given by

$$f_{(1,\ldots,\check{k})}^{\check{k}}(\mathbf{x} \,|\, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{(\check{k}/2)}|\boldsymbol{\Sigma}|^{(1/2)}} \cdot \exp\left(-\frac{1}{2}\left(\mathbf{x}-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)\right) \textit{ for } \mathbf{x} \in \mathbb{R}^{\check{k}}, \tag{3.41}$$

where $\boldsymbol{\mu} \in \mathbb{R}^{\check{k}}$ is the $\check{k}$-dimensional mean and $\boldsymbol{\Sigma} \in \mathbb{R}^{\check{k}\times\check{k}}$ the variance covariance matrix and $|\boldsymbol{\Sigma}|$ its determinant. $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1\check{k}} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2\check{k}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{\check{k}}^2 \end{pmatrix}. \tag{3.42}$$

According to (3.23) and (3.39), the parameter vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{\check{k}})$ and the matrix $\boldsymbol{\Sigma}$ can be estimated by

$$\hat{\boldsymbol{\mu}} = (\bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_{\check{k}}) \;\text{ and }\; \hat{\boldsymbol{\Sigma}} = \begin{pmatrix} s_1^2 & s_{12} & \ldots & s_{1\check{k}} \\ s_{21} & s_2^2 & \ldots & s_{2\check{k}} \\ \vdots & \vdots & \ddots & \vdots \\ s_{21} & s_2^2 & \ldots & s_{\check{k}}^2 \end{pmatrix}, \tag{3.43}$$

for random random samples $\mathbf{x_1}, \ldots, \mathbf{x_{\check{k}}} \in \mathbb{R}^{\check{N}}$ and $\mathbf{x}_{\check{j}} = \left\{x_{1\check{j}}, \ldots, x_{\check{N}\check{j}}\right\}$ for $\check{j} = 1, \ldots, \check{k}$,

Example 3.8. (*Estimation of a Multivariate Normal Distribution*)

Let X_1 and X_2 denote the unknown offset and the gain error of a specific temperature sensor. Both errors shall be assessed in regard of the random samples $\mathbf{x_1} = \{x_{1,1}, \ldots, x_{20,1}\}$ and $\mathbf{x_2} = \{x_{12}, \ldots, x_{20,2}\}$, determined for $\check{N} = 20$ sensors. The boxplots and the correlation coefficient $r_{12} = r_{21} = +0.753$ of both random samples, which are illustrated in figure 3.17, indicate the need for a multivariate normal distribution. Thus, a random vector $\mathbf{X} = (X_1, X_2) \sim N^2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is considered, whose distribution parameters are estimated by (3.43). The estimates are

$$\hat{\boldsymbol{\mu}} = (0.06, 1.00) \ \ and \ \ \hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.2781 & 0.0033 \\ 0.0033 & 7.15e-05 \end{pmatrix}. \tag{3.44}$$
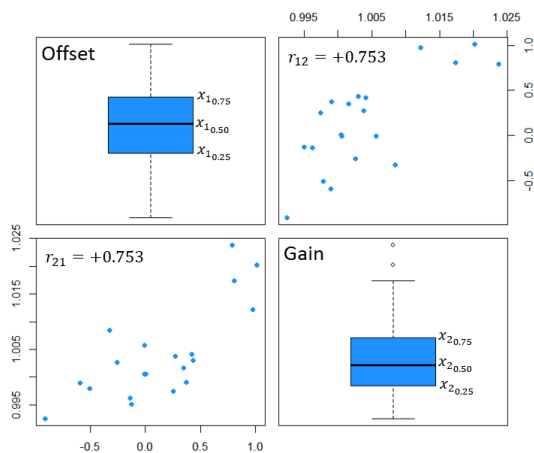
Figure 3.17: Conditional Random Variables: Multivariate Normal Data
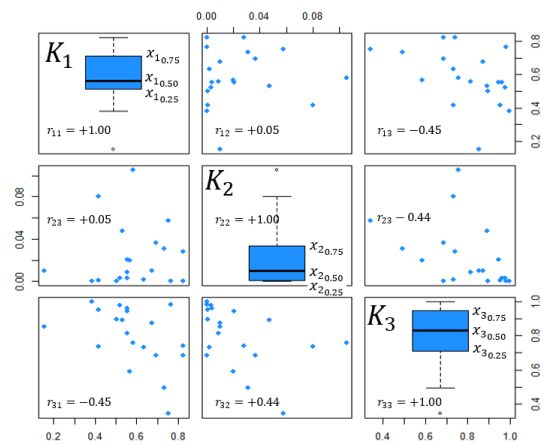
Figure 3.18: Conditional Random Variables: Multivariate Skewed Data

As shown above, the estimation of a multivariate normal distribution based on correlated random samples $\mathbf{x}_1, \ldots, \mathbf{x}_{\check{k}} \in \mathbb{R}^{\check{N}}$ is a straightforward procedure. Still, if two or more correlated random samples are obviously skewed, fitting a multivariate normal distribution is not an appropriate solution anymore. Furthermore, a natural extension of the univariate Beta distribution does not exist for two or more dimensions (cf. Johnson and Kotz [1976]). Although the multivariate *Dirichlet distribution* is a multivariate extension of the Beta distribution, it is not useful for arbitrary data. In fact, the Dirichlet distribution is mainly applied to model rival events (cf. Kotz et al. [2000]). According to that, if two or more correlated random samples are obviously skewed, a nonparametric density can be estimated.

In contrast to the parametric ideology, the nonparametric density estimation approach distances itself from any heuristics and dispenses with any assumptions about a possible shape of the unknown distribution F. A probably well known example of how to nonparametrically estimate the unknown density of available measurement data is the application of a histogram. As already shown for the parametric density estimation approach, the

histogram is rather useful for the instant visual assessment of the data than for estimating the unknown density. The histogram mainly suffers from the discontinuities at the boundary bins, which do not represent the true shape of a continuous distribution. Non-parametric but smooth density estimates are given by *kernel-based* or *nearest-neighbor methods*.

Given that $\mathbf{X} \sim F^{\check{k}}$, the probability $P = \mathbb{P}\left(x \in B \subseteq \mathbb{R}^{\check{k}}\right)$ can be written as

$$P = \int_B f^{\check{k}}(\mathbf{x})\,d\mathbf{x}, \tag{3.45}$$

where $f^{\check{k}}$ is the density function of the unknown distribution $F^{\check{k}}$. Furthermore, let $\mathbf{x}_1, \ldots, \mathbf{x}_{\check{N}}$ denote random samples that were drawn from $F^{\check{k}}$, where $\mathbf{x}_{\check{n}} = \{x_{\check{n}1}, \ldots, x_{\check{n}\check{k}}\}$ for $\check{n} = 1, \ldots, \check{N}$. Then, the probability that $\nu \leq \check{N}$ random samples realize in region $B \subseteq \mathbb{R}^{\check{k}}$ can be calculated by the Binomial distribution as follows.

$$\mathbb{P}\left(\nu \mid \check{N}, P\right) = \binom{\check{N}}{\nu} \cdot P^{\nu} \cdot (1-p)^{\check{N}-\nu}, \tag{3.46}$$

where $\nu := \sum_{\check{n}=1}^{\check{N}} \chi_B(\mathbf{x}_{\check{n}})$ is the sum of indicator functions $\chi_B(\mathbf{x}_{\check{n}}) = \begin{cases} 1 & \text{if } \mathbf{x}_{\check{n}} \in B \\ 0 & \text{if } \mathbf{x}_{\check{n}} \notin B \end{cases}$.

For the Binomial distribution it holds that

$$\mathbb{E}\left(\frac{\nu}{\check{N}}\right) = P \text{ and } VAR\left(\frac{\nu}{\check{N}}\right) = \frac{P \cdot (1-P)}{\check{N}}. \tag{3.47}$$

Hence,

$$\nu/\check{N} \tag{3.48}$$

is a consistent estimator of P, whose distribution is sharply peaked if $\check{N}$ is large (variance goes to zero). For a given sample size $\check{N}$, the accuracy of approximation (3.48) demands a large region B (compared to the value of the density above) so that sufficiently many vectors realize therein. However, if B is sufficiently small and the density function f is continuous (or at least to some extend continuous), the integral in equation (3.45) can be approximated by

$$\int_B f^{\check{k}}(\mathbf{x})\,d\mathbf{x} \approx f^{\check{k}}(\mathbf{x}) \cdot \lambda_{\check{k}}(B), \tag{3.49}$$

where $\lambda_{\check{k}}(B)$ is the $\check{k}$-dimensional *Lebesgue measure* of B. With equations (3.45), (3.48) and (3.49) the unknown density $f^{\check{k}}$ can be finally estimated by

$$f^{\check{k}} \approx \frac{\nu}{\check{N} \cdot \lambda_{\check{k}}(B)}, \tag{3.50}$$

where its validity demanded two contradictory assumptions about the size of region B. Thus, a decision has to be made on how to exploit approximation (3.50). Either it is possible to fix the number of points ν and determine the value of $\lambda_{\check{k}}(B)$ from the sample data, which results in the *nearest-neighbor method*, or it is possible to choose $\lambda_{\check{k}}(B)$ and determine ν from the sample data, giving rise to the *kernel-based method*. Duda and Hart [2000] showed that both, the nearest-neighbor and the kernel-based estimation

method, converge to the true density $f^{\check{k}}$, if $\check{N} \to \infty$. The nearest-neighbor method will not be considered further in this context, because the resulting approximation is not a true density function since its integral over all the sample space diverges. The reader is referred to Bishop [1995] for detailed information about the nearest-neighbor method. For the kernel-based estimation method the region B is defined as a hypercube of side h in $\check{k}$ dimensions centered at $\mathbf{x} \in \mathbb{R}^{\check{k}}$. With the kernel

$$H(\mathbf{u}) = \begin{cases} 1 & \text{if } |u_j| < 1/2 \ \forall \check{j} = 1, \dots, \check{k} \\ 0 & \text{else} \end{cases}, \tag{3.51}$$

that is the *Parzen Window*, the statistic

$$\nu = \sum_{\check{n}=1}^{\check{N}} H\left(\frac{\mathbf{x} - \mathbf{x}_{\check{n}}}{h}\right) \tag{3.52}$$

counts all random samples inside of hypercube B. Approximation (3.50) can be rewritten as

$$f^{\check{k}}(\mathbf{x}) \approx \frac{1}{\check{N} \cdot h^{\check{k}}} \cdot \sum_{\check{n}=1}^{\check{N}} H\left(\frac{\mathbf{x} - \mathbf{x}_{\check{n}}}{h}\right). \tag{3.53}$$

After discretizing the domain of the density into support vectors $\mathbf{x}$, $f^{\check{k}}$ can be estimated by successively applying approximation (3.53). Yet, the density estimator provides discontinuities, which can be however eliminated by exchanging the kernel of equation (3.51) with the Gaussian kernel

$$H(\mathbf{x}_{\check{n}}) = \frac{1}{\check{N}} \cdot \sum_{\check{n}=1}^{\check{N}} \frac{1}{(2\pi)^{\check{k}/2}} \cdot \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_{\check{n}}\|^2}{2h^2}\right) \tag{3.54}$$

so that

$$f^{\check{k}}(\mathbf{x}) \approx \frac{1}{\check{N}} \cdot \sum_{\check{n}=1}^{\check{N}} \frac{1}{(2\pi h^2)^{\check{k}/2}} \cdot \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_n\|^2}{2h^2}\right) = \frac{1}{\check{N}} \cdot \sum_{\check{n}=1}^{\check{N}} f_{\check{n}}^{\check{k}}(\mathbf{x}). \tag{3.55}$$

The side length h of the hypercube B, which is often denoted as *bandwidth*, corresponds to the constant standard deviation of the normal distributions $f_{\check{n}} :\equiv N(\mathbf{x}_{\check{n}}, \mathbb{I}_{\check{k}} h^2)$ for $\check{n} = 1, \dots, \check{N}$. As highlighted in figure 3.19, approximation (3.55) adds up the contributions of $\check{N}$ normal distributions, spanned by $\mathbf{x}_1, \dots, \mathbf{x}_{\check{N}}$, for $\mathbf{x} \in \mathbb{R}^{\check{k}}$, and it standardizes the resulting value by N so that a density estimate is provided. A particular advantage of Gaussian kernels is that an artificial random sample is easily generated by generating random samples of $N(\mathbf{x}_{\check{n}}, \mathbb{I}_{\check{k}} h^2)$. The alternating means $\mathbf{x}_{\check{n}}$ are uniformly drawn (with replacement) from the existing data, that is $\mathbf{x}_1, \dots, \mathbf{x}_{\check{N}}$.

Example 3.9. *(Estimation of a Multivariate Nonparametric Distribution)*

*Manufacturing tolerances in regard of a **D**iesel **O**xidation **C**atalyst (DOC) shall be investigated. On this account, a random sample of $\check{N} = 20$ catalysts is monitored and recorded. The data obtained is used to fit the preexponential constants $K_1 = X_1$ (CO-oxidation), $K_2 = X_2$ (HC-oxidation) and $K_3 = X_3$ (NO-oxidation) of the implemented Arrhenius' equations. The resulting random samples $\mathbf{x}_1 = \{x_{1,1}, \dots, x_{20,1}\}$, $\mathbf{x}_2 = \{x_{1,2}, \dots, x_{20,2}\}$ and*
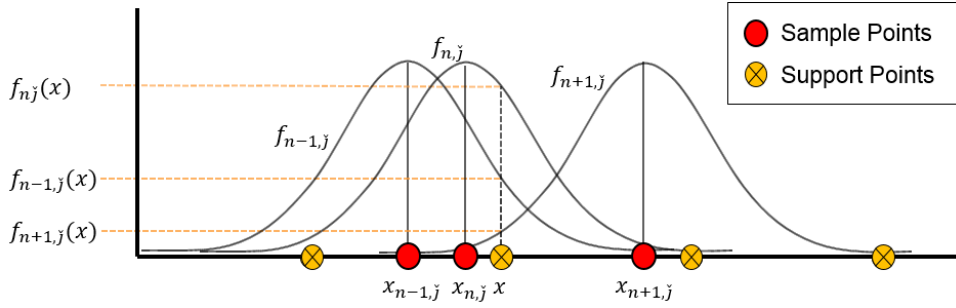
Figure 3.19: Nonparametric Density Estimation

$\mathbf{x}_3 = \{x_{1,3}, \ldots, x_{20,3}\}$ *are skewed and correlated so that a 3-dimensional nonparametric distribution is fitted. The scatterplot-matrix in figure 3.18 illustrates the linearly transformed random samples* $\mathbf{x}_{\check{j}}^{std}$ *(cf. transformation (3.32)) as well as the correlation coefficients* $r_{\check{i}\check{j}}$ *for* $\check{i}, \check{j} = 1, 2, 3$. *The* **R** *package "ks" provides all functions necessary to nonparametrically fit 3-dimensional data with Gaussian kernels. Figure 3.20 illustrates a possibility for visualizing such a nonparametrically estimated density in 3 dimensions. Among others, the optimal bandwidth h can also be determined with that package. Artificial observations can be gathered with that bandwidth, as illustrated by figure 3.21.*
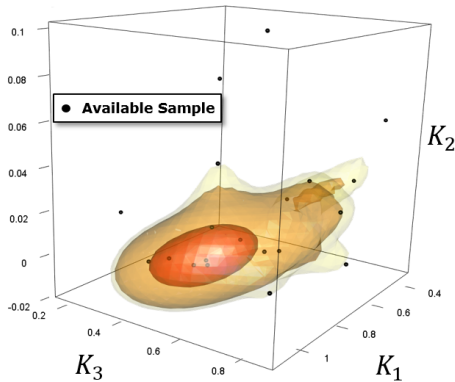
Figure 3.20: Nonparametric Density Estimate for Skewed Multivariate Data ($\check{N} = 20$)
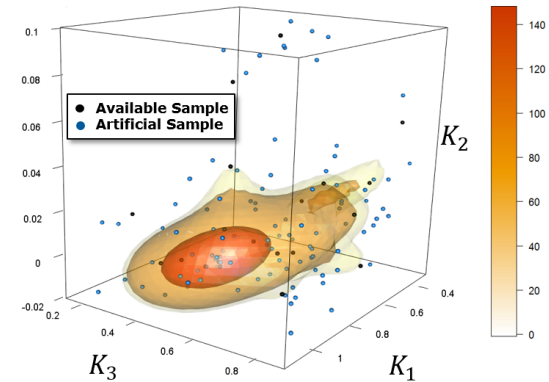
Figure 3.21: Nonparametric Density Estimate: Artificial Random Sample of Size 100

   Manufacturing tolerances of the introduced production and lifetime-oriented process are expressed by the random variables $X_1, \ldots, X_{\check{k}}$. Due to the semi-physical structure of the employed simulation tool MoBEO, these are not necessarily independent. Accordingly, a more general approach is to consider a random vector $\mathbf{X}$. In this thesis, its elements may follow continuous multivariate distributions, as outlined by (3.56).

$$\mathbf{X} = \left(\overbrace{X_1, \ldots,}^{F^1} \overbrace{X_{\check{i}}, X_{\check{i}+1}, \ldots,}^{\sim F^2} \overbrace{X_{\check{j}-1}, X_{\check{j}}, X_{\check{j}+1}, \ldots,}^{\sim F^3} \overbrace{X_{\check{k}}}^{\sim F^1} \right) \sim F^{\check{k}}. \tag{3.56}$$

The density of distribution $F^{\check{k}}$ is determined for the special case of (3.56) by the product of independent terms as follows

$$f^{\check{k}}_{(1,\ldots,\check{k})}(\mathbf{x}) = f^1_1(x_1) \cdot \cdots \cdot f^2_{\check{i},\check{i}+1}(x_{\check{i}}, x_{\check{i}+1}) \cdot \cdots \cdot f^3_{\check{j}-1,\check{j},\check{j}+1}(x_{\check{j}-1}, x_{\check{j}}, x_{\check{j}+1}) \cdot \cdots \cdot f^1_{\check{k}}(x_{\check{k}}). \quad (3.57)$$

In general, the dependent clusters may differ from (3.57) so that the multiplication has to be adapted accordingly. In spite of the generalization to the random vector $\mathbf{X}$, it should be pointed out that most information is gathered by independent random variables. Once the distribution $F^{\check{k}}$ is determined for $\mathbf{X}$, manufacturing tolerances can be simulated by accessing and modifying the right MoBEO parameters in regard of distribution $F^{\check{k}}$ (cf. subsection 3.1.2).

An existent parameterization providing accurate simulation results is the basis for the simulation of manufacturing tolerances. A natural starting point is a parameterization that represents the complete production series in the best possible way. Here, the SPC, introduced in subsection 3.1.1, suggests using a *nominal parameterization*, giving rise to the so called *nominal vehicle*.

Definition 3.18. (*The Nominal Parameterization*)

Given that $X_{\check{j}} \overset{iid}{\sim} F_{\check{j}}$, the nominal engine model environment provides parameters $x_1, \ldots, x_{\check{k}}$ that equal to the median values

$$x_{\check{j}} := m_{\check{j}} = q_{\check{j}0.5} \; for \; \check{j} = 1, \ldots, \check{k}.$$

A MoBEO model comprising a nominal parameterization models an engine environment, which approximately divides the production series into two halves regarding $X_1, \ldots, X_{\check{k}}$. Hence, approximately 50% of the production would be simulated by $x_{\check{j}} \leq m_{\check{j}}$ (and by $x_{\check{j}} > m_{\check{j}}$ vice versa) for all $\check{j}$. Still, the SPC and consequently definition 3.18 presumes independent random variables $X_1, \ldots, X_{\check{k}}$. Regarding the use of MoBEO, definition 3.18 is generalized as follows.

Definition 3.19. (*The Generalized Nominal Parameterization*)

Given that $\mathbf{X} = (X_1, \ldots, X_{\check{k}}) \sim F^{\check{k}}$, the nominal engine model environment provides parameters $x_1, \ldots, x_{\check{k}}$ that satisfy

$$\mathbf{x} := \mathbf{M}_{F^{\check{k}}},$$

where $\mathbf{M}_{F^{\check{k}}}$ denotes the $\check{k}$-dimensional median of distribution $F^{\check{k}}$. The multidimensional median will be defined and discussed in chapter 4.

The facility to work with the nominal engine is a big advantage of the simulation tool MoBEO compared with a real test bed, where an arbitrary engine environment is investigated. Statements from the test bed may be slightly biased or even not representative for the underlying production series. This is especially the case, when the test bed engine environment comprises realizations of decisive random variables that are far away from the relating nominal values, as illustrated by figure 3.22.

A probabilistic modeling of the development stage was given in this section. The factor
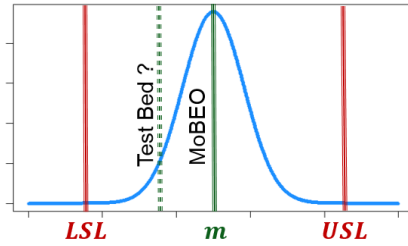
Figure 3.22: Simulation of the Nominal Vehicle

combinations of the open system parameters, represented by uniform random variables $U_1, \ldots, U_{\tilde{k}}$, and the manufacturing tolerances, considered by continuous random variables $X_1, \ldots, X_{\check{k}}$, need to be simulated and investigated. On this account, especially manufacturing tolerances were discussed in detail due to two reasons. Firstly, the number $\check{k}$ is larger in practice than $\tilde{k}$ and, secondly, the distributions regarding manufacturing tolerances are usually not uniform distributions, as assumed for the open system parameters. Subsection 3.1.1 introduced the probabilistic background for manufacturing tolerances. Then, subsection 3.1.2 highlighted the problem that manufacturing tolerance information need to be processed very carefully with regard to the semi-physical simulation tool MoBEO. Subsection 3.1.2 discussed estimation procedures with regard to manufacturing tolerances and introduced the idea of the nominal vehicle.

3.2   Modeling and Implementing the Usage Stage

The variability of the response variables concerning the production is captured by the random variables $X_1, \ldots, X_{\check{k}}$. In contrast, the variability transmitted by the lifetime usage is involved by different usage behavior and various environmental conditions. The method how these relating factors can be modeled and implemented into MoBEO still needs clarification.

Opposite to the factors of the development stage, environmental conditions or usage behavior are not controllable in practice. Hence, a reasonable consideration requires, first of all, representative measurement data that needs to be carefully acquired. Once this is achieved, the usage stage can be involved by conducting simulations under different environmental conditions and usages. Environmental conditions are considered by quantitative features as seasonal ambient temperatures, pressures and humidities under which a vehicle is mainly operated. If environmental conditions are held constant during a simulation, their implementation corresponds to the implementation of manufacturing tolerances. In spite of environmental conditions, usage behavior is considered as a discrete feature that can realize in so called *usage profiles*, denoting portions of (almost) the total usage in the field, like *city-driving* or *mountain-driving* for example.

3.2.1   Environmental Conditions

Environmental conditions, such as ambient temperature, pressure or humidity, may have a significant impact on the performance of a vehicle. From the perspective of the modeled

engine environment, these factors mainly affect the intake air path or the combustion of the engine. In statistics, environmental factors denote continuous correlated features that can be modeled by continuous random vector $\mathbf{V} \sim F^3$. Once, the three-dimensional distribution F^3 is estimated, these can be simulated by MoBEO, as manufacturing tolerances. During the usage stage, a vehicle is exposed to various environmental conditions. Primarily, these depend on the time of the year and the location, where the vehicle is mainly operated. Although exceptional ambient conditions are corrected by several control strategies, these corrections are only effective to a certain extent. Especially beyond the *normal conditions of use* (cf. subsection A.2), the engine responsiveness, emissions or fuel consumption can be noticeably affected. This is because environmental conditions affect the intake air path, cooling systems, the combustion, heat transfers, oil viscosities or the catalyst light-off behavior. For all that, environmental conditions play a crucial role for the introduced production and lifetime-oriented process. Their implementation is realized by setting the corresponding environmental MoBEO parameters on demanded values before starting a simulation. A representative consideration of environmental conditions requires determining their underlying distributions.

Especially in a country, like Austria, vehicles are exposed to a broad range of temperature and pressure levels over the whole year. The annual average temperature, for example, varies between $-9°C$ at the summit of the Großglockner ($3798\,m$ or $\approx 63.50\,kPa$) and $+11°C$ in Vienna ($157\,m$ or $\approx 98.03\,kPa$) (cf. Hiebl et al. [2010]). In Graz (353 or $\approx 97.72\,kPa$), the second largest city of Austria, the average temperature is distributed between $-3.7°C$ and $+25.3°C$ over the year. Comparable to Austria, Europe has a significant environmental spread in terms of ambient temperature and pressure. From sea level to the highest road in Europe, the *Col de la Bonette* ($2715\,m$ or $\approx 72.55\,kPa$) in the French Alps, from average temperatures of $10.0°C$ in Stockholm to $22.3°C$ in Athens, Europe exhibits a broad spectrum of environmental conditions. The internet offers plenty of information concerning global climate data. The page www.wetterkontor.de, for example, provides seasonal temperature and humidity data for many countries and their cities all over the globe. Dependent on the application of the vehicle, the available data shall be combined to derive representative environmental profiles, as presented in the following example.

Example 3.10. (*Generation of Environmental Profiles*)

During the production and lifetime-oriented development, the usage stage of a specific compact car shall be modeled for Europe. Among others, this concerns the environmental conditions under which the sold vehicles are mainly operated. On this account, the manufacturer provides the automotive consulting firm the predicted sales figures of the most relevant countries, contributing to approximately 90% of the European sales. These sales figures are outlined in table 3.3. Assuming that compact cars are mainly driven in urban areas, it makes sense to focus the largest metropolitan areas of the countries given. Here, table 3.3 also provides the lowest and highest seasonal ambient temperatures in degree Celsius, and table 3.4 outlines the corresponding absolute humidities in g/kg. The different sales figures propose a corresponding weighting of the measurement data given. Therefore, the average temperatures (between low and high) and humidities are replicated in respect of the sold vehicles. For this example, replicates are generated until $\check{N} = 100$ observations are available. The resulting 2-dimensional observations are given by figure 3.23. The correlation coefficient clearly indicates the need for a 2-dimensional

Country	Predicted Sales	Metropolitan Region	[1961-1990]	Jan		Feb		Mar		Apr		May		Jun		Jul		Aug		Sep		Oct		Nov		Dec	
				low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high	low	high
Germany	33.5%	Rhine-Ruhr	Ambient Temperature [°C] www.wetterkontor.de	-1.0	4.0	-1.0	5.0	2.0	9.0	4.0	13.0	8.0	18.0	11.0	21.0	13.0	22.0	13.0	22.0	10.0	19.0	7.0	15.0	3.0	9.0	1.0	5.0
		Munich		-5.1	1.6	-4.0	3.6	-0.8	8.1	2.6	12.6	6.8	17.4	10.2	20.5	12.1	22.8	11.8	22.3	8.9	19.1	4.4	13.6	-0.1	6.9	-3.7	2.6
Russia	15.3%	Moscow		-12.7	-6.2	-11.6	-3.9	-5.9	2.4	1.8	10.6	7.6	18.6	11.4	22.4	13.1	23.8	11.7	22.0	7.0	15.8	2.1	8.4	-3.4	1.3	-8.9	-3.4
Italy	14.0%	Milan		-0.6	4.6	0.9	7.6	5.0	13.0	8.7	18.2	13.0	23.2	16.7	27.8	19.1	30.4	18.6	29.2	15.4	24.3	10.1	17.3	5.0	10.3	1.1	5.8
		Rome		3.7	12.9	4.4	13.7	5.8	15.3	8.3	18.0	11.9	22.0	15.6	25.6	18.2	28.6	18.4	28.7	15.8	26.0	12.0	22.0	8.1	17.2	5.1	13.9
France	10.8%	Paris		0.9	6.0	1.3	7.4	3.6	12.2	6.3	15.8	9.5	19.7	12.7	22.9	14.5	24.6	14.3	24.0	11.9	21.1	7.9	15.6	4.5	10.0	2.0	6.6
Great Britain	7.0%	London		2.2	6.4	2.2	7.1	3.3	10.1	5.5	13.3	8.2	16.9	11.6	20.3	13.5	21.8	13.2	21.4	11.3	18.5	7.9	14.2	5.3	10.1	3.5	7.4
Spain	6.2%	Madrid		1.5	8.5	2.2	11	5.2	14.9	7.4	18.4	10.2	21.2	14.6	26.9	17.4	30.8	17.1	29.5	14.1	25.0	9.5	18.5	5.3	12.8	2.2	8.8
Turkey	2.9%	Istanbul		2.9	8.8	3.2	9.4	4.3	11.6	8.0	16.7	12.0	21.5	16.1	26.2	18.5	28.2	18.5	28.1	15.5	25.0	11.8	19.8	8.5	15.4	5.4	11.2

Table 3.3: Environmental Profiles: Ambient Temperature

Country	Predicted Sales	Metropolitan Region	[1961-1990]	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Germany	33.5%	Rhine-Ruhr	Absolute Humidity [g/kg] www.wetterkontor.de	3.7	3.6	4.4	5.1	6.7	8.5	9.3	9.4	8.3	6.8	5.0	4.1
		Munich		2.8	3.2	3.9	4.8	6.6	8.1	9.3	9.3	7.9	6.0	4.3	3.2
Russia	15.3%	Moscow		1.6	1.8	2.7	4.3	6.1	8.2	9.8	9.4	6.9	4.7	3.1	2.1
Italy	14.0%	Milan		3.8	4.0	4.8	6.0	7.9	9.9	11	11.2	9.8	7.6	5.5	4.2
		Rome		5.2	5.5	6.1	7.2	9.2	11.3	13.3	13.6	11.9	9.4	7.2	5.8
France	10.8%	Paris		4.3	4.3	5.1	5.8	7.5	9.5	10.6	10.5	9.5	7.5	5.6	4.6
Great Britain	7.0%	London		4.6	4.5	4.9	5.5	6.7	8.1	9.3	9.5	8.6	7.1	5.9	5.1
Spain	6.2%	Madrid		4.3	4.3	4.8	5.5	6.1	7.6	7.6	7.6	7.5	6.7	5.3	4.4
Turkey	2.9%	Istanbul		4.7	4.8	5.2	6.7	9.0	11.4	12.9	12.8	11.2	8.9	7.1	5.6

Table 3.4: Environmental Profiles: Ambient Humidity

distribution. This should be estimated by a nonparametric Gaussian Kernel approach, as outlined in subsection 3.1.3. Hence, ambient temperature and humidity is modeled by two stochastically dependent random variables V_1 and V_2. The resulting density and a corresponding artificial random sample with 200 observations is illustrated in figure 3.24. In regard of the ambient pressure, the manufacturer proposes the use of an independent Beta distributed random variable $V_3 \sim Beta(\boldsymbol{\theta})$ with parameters $\boldsymbol{\theta} = (4.5, 1.0)$ on $[0.7, 1.0]$ bar. Hence, representative environmental conditions of the objective compact car are simulated by setting the respective MoBEO parameters on realized values of the random vector

$$\mathbf{V} = \left(\overbrace{(V_1, V_2)}^{temp., \ hum.}, \overbrace{V_3}^{pres.} \right) \sim F^3. \tag{3.58}$$

So far, the environmental conditions of the usage stage can be simulated for ambient temperature, pressure and humidity by MoBEO. Just as for manufacturing tolerances, these are set once at the beginning of a simulation. Although, environmental conditions may change while the vehicle is operated, the simplification to constant environmental conditions is a plausible measure with regard to average statements. The environmental profiles to be generated have to be determined with the internal sales figures of the manufacturer and in dependence of the vehicle application. The operation radius of a compact car will certainly be smaller on average than the one of a truck. Beside environmental conditions, the usage behavior of potential end-users shall be modeled.
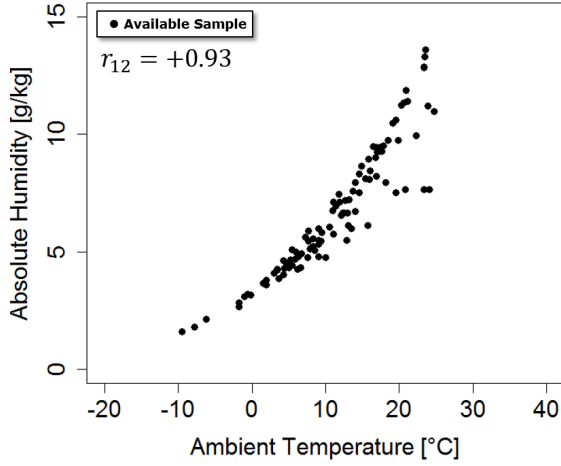
Figure 3.23: Environmental Profiles: Ambient Temperature vs. Humidity ($\check{N} = 100$)
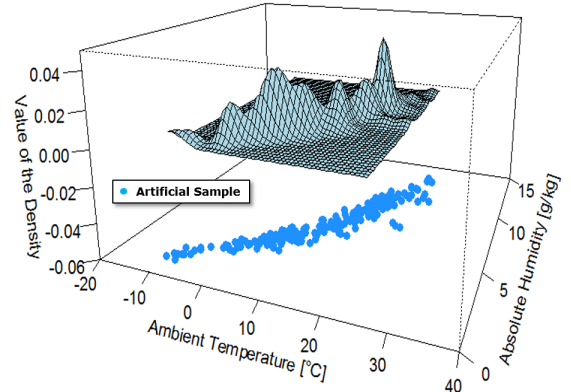
Figure 3.24: Environmental Profiles: Artificial Random Sample of Size 200

3.2.2 Usage Behavior

Modeling and implementing the usage stage requires, beside the consideration of environmental conditions, the analysis of the usage behavior of prospective end-users. So called *usage profiles* are characterized in order to enable a representative coverage of possible driving patterns. The reader is referred to Kirschbaum [2013] for a detailed explanation about how usage behavior is investigated. In this subsection, an overview of the approaches presented therein is given.

Modeling the usage behavior of prospective end-users requires finding out which usage related features $\Xi_1, \ldots, \Xi_{\check{m}}$ are the most decisive for delimitable usage profiles $\{z_1, \ldots, z_{\check{k}}\}$. While such usage related features might be for example the engine speed, torque or vehicle speed, the usage profiles, such as city-driving, highway-driving, or mountain-driving, denote the data origin of the usage related features considered. If measurement data $\boldsymbol{\xi}_{1_{\check{j}}}, \ldots, \boldsymbol{\xi}_{\check{m}_{\check{j}}}$ of features $\Xi_{1_{\check{j}}}, \ldots, \Xi_{\check{m}_{\check{j}}}$ is multiple times available for usage profile $z_{\check{j}}$, $\check{j} = 1, \ldots, \check{k}$, the observations can be summarized by a matrix

$$\boldsymbol{\Psi}_{\check{j}} = \begin{pmatrix} \boldsymbol{\xi}_{(11)_{\check{j}}} & \boldsymbol{\xi}_{(12)_{\check{j}}} & \cdots & \boldsymbol{\xi}_{(1\check{m})_{\check{j}}} \\ \boldsymbol{\xi}_{(21)_{\check{j}}} & \boldsymbol{\xi}_{(22)_{\check{j}}} & \cdots & \boldsymbol{\xi}_{(2\check{m})_{\check{j}}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\xi}_{(\check{N}1)_{\check{j}}} & \boldsymbol{\xi}_{(\check{N}2)_{\check{j}}} & \cdots & \boldsymbol{\xi}_{(\check{N}\check{m})_{\check{j}}} \end{pmatrix} \tag{3.59}$$

where $\boldsymbol{\xi}_{(\check{n}\check{i})_{\check{j}}} \in \mathbb{R}^{t_{\check{n}_{\check{j}}}}$ for $\check{n} = 1, \ldots, \check{N}$ and $\check{i} = 1, \ldots, \check{m}$. Thus, the considered features are neither necessarily the same nor of the same number for all usage profiles. Moreover, the lengths of the observed data may differ from row to row and from usage profile to usage profile so that each entry of $\boldsymbol{\Psi}_{\check{j}}$ is a time sequence of $t_{\check{n}_{\check{j}}}$ observations. As Kirschbaum [2013] outlines, it is proposed to use only measurement data that complies with $t_{\check{n}_{\check{j}}} > 4h$. Moreover, observations $\boldsymbol{\xi}_{(\check{n}\check{i})_{\check{j}}} \in \mathbb{R}^{t_{\check{n}_{\check{j}}}}$ are frequently fragmentary and may

contain inappropriate measurement errors so that $\mathbf{\Psi}_{\breve{j}}$ has to be considered in a more robust manner. This is overcome by determining numerous statistics, like empirical quantiles, the **Inter**Q**uartile** **R**ange** (IQR) or the **Median** **A**bsolute** **D**eviation** (MAD) for every feature $\Xi_{i_{\breve{j}}}$. Each row

$$\left(\boldsymbol{\xi}_{(11)_{\breve{j}}}, \boldsymbol{\xi}_{(12)_{\breve{j}}}, \ldots, \boldsymbol{\xi}_{(1\breve{m})_{\breve{j}}}\right) \tag{3.60}$$

of $\mathbf{\Psi}_{\breve{j}}$ is transformed to

$$\left(\xi_{(\breve{n}1)_{\breve{j}}}, \xi_{(\breve{n}2)_{\breve{j}}}, \ldots, \xi_{(\breve{n}\breve{m}')_{\breve{j}}}\right), \tag{3.61}$$

by this robustness measure, so that $\xi_{(\breve{n}\breve{i}')_{\breve{j}}} \in \mathbb{R}$ for $\breve{i}' = 1, \ldots, \breve{m}'$ and $\breve{m}' > \breve{m}$. Thus,

$$\mathbf{\Psi}_{\breve{j}}^{\text{robust}} = \begin{pmatrix} \xi_{(11)_{\breve{j}}} & \xi_{(12)_{\breve{j}}} & \cdots & \xi_{(1\breve{m}')_{\breve{j}}} \\ \xi_{(21)_{\breve{j}}} & \xi_{(22)_{\breve{j}}} & \cdots & \xi_{(2\breve{m}')_{\breve{j}}} \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{(\breve{N}1)_{\breve{j}}} & \xi_{(\breve{N}2)_{\breve{j}}} & \cdots & \xi_{(\breve{N}\breve{m}')_{\breve{j}}} \end{pmatrix} \tag{3.62}$$

is obtained, whose column vectors denote observations of the robust features $\Xi_{1_{\breve{j}}}^{\text{robust}}, \ldots, \Xi_{\breve{m}'_{\breve{j}}}^{\text{robust}}$. The objective is to find the most decisive robust features for all available usage profiles $z_{\breve{j}}$, $\breve{j} = 1, \ldots, \breve{k}$. Kirschbaum [2013] achieves this by conducting a *robust principal component analysis* (Robust PCA), as described by Croux et al. [2013]. Robust PCA is capable of identifying those features, which possess 70% to 90% of total variability possessed by all features. Conducting the robust PCA on $\mathbf{\Psi}_{\breve{j}}^{\text{robust}}$ generally yields a few decisive robust features, characterizing the investigated usage profile. The results show that load and dynamic features are often outstanding, as illustrated for two dimensions by figure 3.25. It is clearly visible that the usage profile *city driving*, for instance, is almost completely separated from the other usage profiles, because it takes place in low load but high dynamic areas. Indeed, every illustrated point corresponds to a certain time sequence of usage related features $\Xi_{1_{\breve{j}}}, \ldots, \Xi_{\breve{m}_{\breve{j}}}$. When using the simulation tool MoBEO, only the usage related features engine speed and engine torque can be processed. In accordance to example 3.10, it could be the target for MoBEO to simulate the city driving usage profile. Then, the city driving usage profile needs to be characterized by one or more gray triangles. The probably most simple approach is to randomly choose one of the gray triangles, and simply look after the corresponding engine speed and engine torque trace. Nevertheless, it might be of interest to simulate the data center, especially then when the simulation effort is limited. The data center is illustrated by figure 3.26 for the usage profiles city-driving, highway-driving, or mountain-driving. If every usage profile has to be simulated multiple times, the simulations could correspond to the multi-dimensional quantiles, which are introduced in section 4.2. On effect of this measure, the multidimensional distribution of a usage profile is well covered. In any case, once it is clear on how to simulate individual usage profiles, the distribution of the usage profiles needs to be considered. On this account, it is assumed that the general vehicle usage is a mixture of usage profiles $\{z_1, \ldots, z_{\breve{k}}\}$. Consequently, from the probabilistic point of view usage profiles are modeled by a discrete multinominal distributed random variable $\mathbf{Z} = (Z_1, \ldots, Z_{\breve{k}})$, whose realization values correspond vectors $\mathbf{z} = (z_1, \ldots, z_{\breve{k}})$. The multinomial distribution
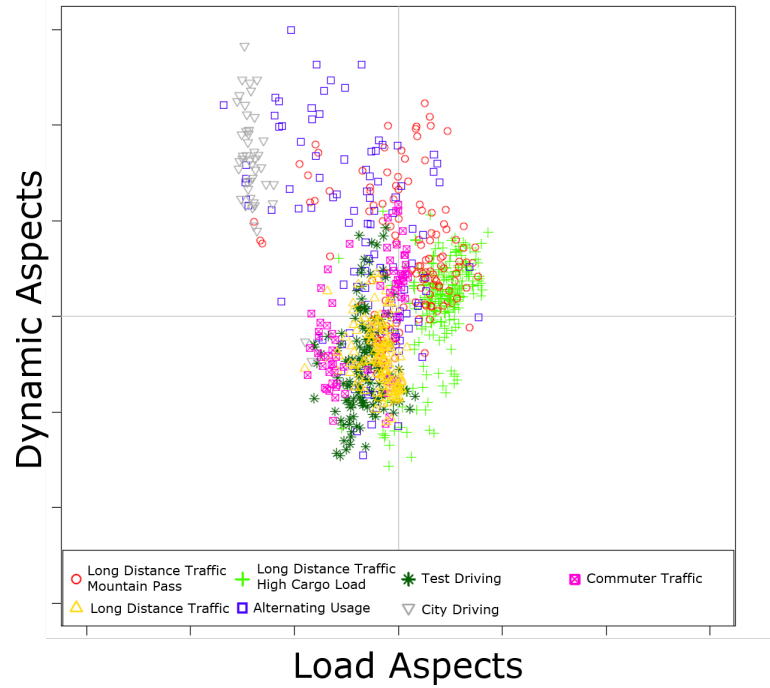
Figure 3.25: Characterization of Usage Profiles

is defined as

$$\mathbb{P}\left(Z_1 = z_1, \dots, Z_2 = z_2, \dots, Z_{\check{k}} = z_{\check{k}}\right) = \frac{\check{N}!}{z_1! \cdot z_2! \cdots z_{\check{k}}!} \cdot p_1^{z_1} \cdot p_2^{z_2} \cdots p_{\check{k}}^{z_{\check{k}}}, \qquad (3.63)$$

where $p_1 + p_2 + \cdots + p_{\check{k}} = 1$, describe the theoretical portion of each usage profile with regard to the total use (or a majority thereof, as illustrated by figure 3.26 for 90% of the total use). These portions shall be obtained with the aid of the manufacturer. For given a given *usage behavior* $\mathbf{z} = (z_1, \dots, z_{\check{k}})$, the simulations shall be conducted in accordance to the empirical probabilities $\hat{p}_1, \dots, \hat{p}_{\check{k}}$. Thus, if city-driving is three times as probable as highway-driving for the usage stage of a vehicle, then city-driving has to be approximately simulated three times more often than highway-driving.

The European legislation, for example, provides the *normal conditions of use* and transient emission cycles[4] like the *NEDC*[5] for passenger cars, the *WHTC*[6] or the *NRTC*[7] for heavy-duty applications in order to cover the feasible emission spectrum during the usage stage of a vehicle. In contrast to legislative requirements, the introduced development and lifetime-oriented development process considers the usage stage without necessarily focusing on emissions. The general view on the usage stage provides the possibility to investigate the impact of the development stage on the usage stage. Beside the exhaust behavior, this especially concerns the associated aging of the vehicle, thereby making possible geographical influenced warranty statements for example. The joint consideration of

---

[4]cf. subsection A.2.1
[5]"New European Driving Cycle", cf. Delphi [2012]
[6]"World Harmonized Transient Cycle", cf. Delphi [2012]
[7]"Nonroad Transient Cycle", cf. Delphi [2012]
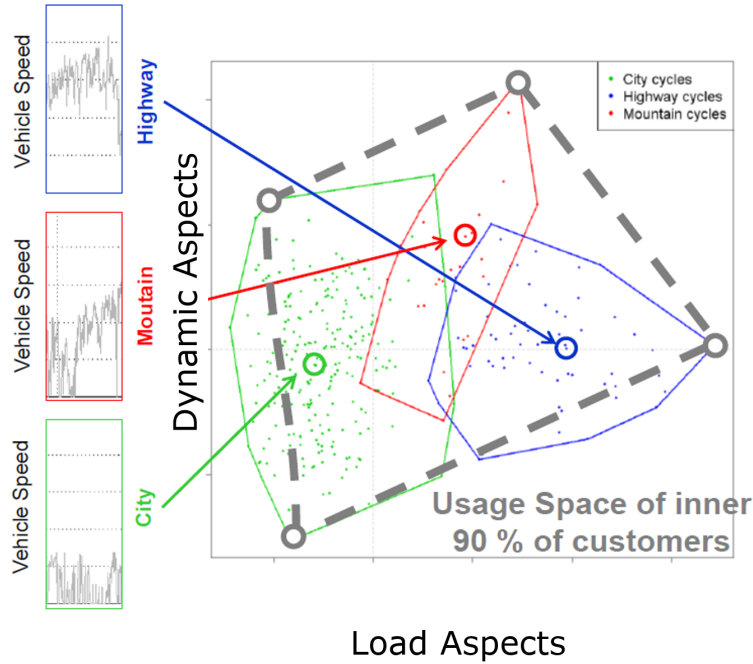
Figure 3.26: Identification of Usage Profiles

the development and usage stage is expressed by the feature space that will be introduced in the next section.

3.3   The Feature Space

The launch of the introduced production and lifetime-oriented process means starting to investigate factor combinations in terms of the objective engine environment. Due to the last sections it should be clear that the factors denote random variables, whose realizations are generally subject to limited domains. The joint consideration of all factors yields therefore a multidimensional random vector that is subject to a multidimensional domain, that is the so called *feature space*. This section will deal with the definition of the feature space and the modeling of its underlying multidimensional distribution.

In section 3.1 and 3.2 the factors of the development and usage stage have been discussed in detail. Table 3.5 shortly recapitulates the findings concerning the controllable and uncontrollable factors of the RPD. The joint consideration of all factors yields the random vector

$$\mathbf{X} = (U_1, \ldots, U_{\tilde{k}}, X_1, \ldots, X_{\check{k}}, V_1, V_2, V_3, \mathbf{Z}) \sim F^k. \tag{3.64}$$

Every realization $\mathbf{x}$ of $\mathbf{X}$ shall be denoted as *experiment*, meaning that $\mathbf{x}$ corresponds to one simulation run with MoBEO. Under the agreement that $k = \tilde{k} + \check{k} + \check{k} + 3$, every experiment corresponds to a k-dimensional vector.

Definition 3.20. (*Feature Space*)

The set of all experiments $\mathbf{x}$ is defined as feature space $\mathbb{X}^k$. Given that $\Omega_1, \ldots, \Omega_k$ denote
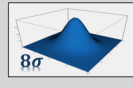
	Considered Factors		Features	Distribution Types	Impact on MoBEO
Development Stage	**Open System Parameters**		$U_1, \dots, U_{\tilde{k}}$	**Discrete** Uniform	**Model- or Parameter Modification**
				Continuous Uniform	**Parameter Modification**
	Manufacturing Tolerances	8σ	$X_1, \dots, X_{\check{k}}$	**Continuous** Normal, Beta, Non-Parametric	**Parameter Modification**
Usage Stage	**Environmental Conditions**		V_1, V_2, V_3	**Continuous** Normal, Beta, Non-Parametric	**Parameter Modification** (Ambient Temperature, Humidity and Pressure)
	Usage Behavior		Z	**Discrete** Multinomial Distribution	**Usage Profile to be Simulated by Engine Speed and Torque**

Table 3.5: Production and Lifetime-Oriented Development: Factor Overview

the marginal domains of $X_1, \dots, X_k$, the Cartesian product of the marginal domains spans the feature space. Thus,

$$\mathbb{X}^k = \Omega_1 \times \cdots \times \Omega_k \subset \mathbb{R}^k. \tag{3.65}$$

Dependent on whether or not $\mathbf{X}$ contains discrete random variables, $\mathbb{X}^k$ is a discrete space or a continuous space, respectively.

In case of open system parameters, the domains correspond to all feasible quality standards that have to be investigated. Thus, if the performances of a vanadium SCR catalyst and an iron-zeolite SCR catalyst shall be compared, the domain of this discrete consideration $U_{\tilde{j}}$ is given by $\Omega_{\tilde{j}} = \{1, 2\}$. In contrast, if the optimal light-off temperature $U_{\tilde{j}}$ of an iron-zeolite SCR catalyst has to be detected, the respective domain is an interval (usually $\Omega_{\tilde{j}} = [373.15\,K, 673.15\,K]$). With regard to manufacturing tolerances and barely capable production processes, setting $\Omega_{\check{j}} = T_{\check{j}} = [LSL_{\check{j}}, USL_{\check{j}}]$ would be a reasonable choice for all $\check{j} = 1, \dots, \check{k}$. Concerning environmental conditions, the domains shall be chosen such that the measurement data is well covered. With respect to the usage behavior the domain of random variable Z equals to $\Omega = \{1, \dots, \check{k}\}$.
Spanning a feature space shall be always conducted with regard to the response variables considered. This is because each additional dimension of $\mathbb{X}^k$ causes the investigation effort to exponentially grow. The reader is referred to Bishop [1995], where this problem is well discussed under the synonym "*curse of dimensionality*". For that reason, it is always proposed to preselect the considered factors in respect of response variables of interest. In this thesis, these are denoted by $Y_1, \dots, Y_r$. Especially, the investigation of particular manufacturing tolerances should be synchronized with expert knowledge. It makes no sense to include manufacturing tolerances of components that have almost no influence on the investigated response variable(s). As a result to that restriction, $\mathbb{X}^k$ can be seen as a superset of all de facto produced vehicles and their way of usage. Indeed, every experiment $\mathbf{x} \in \mathbb{X}^k$ correspond to a particular vehicle assembly that is exposed to a certain kind of usage. With the assumptions made in section 2.2, the *cdf* of the underlying

feature space distribution F^k corresponds to

$$F^k(\mathbf{x}) = \left(\prod_{\tilde{j}=1}^{\tilde{k}} \mathbb{P}\left(U_{\tilde{j}} \le u_{\tilde{j}}\right)\right) \cdot \mathbb{P}\left(\mathbf{X}_{\check{k}} \le \mathbf{x}_{\check{k}}\right) \cdot \mathbb{P}\left(\mathbf{V} \le \mathbf{v}\right) \cdot \mathbb{P}\left(\mathbf{Z} \le \mathbf{z}\right) \ \text{ for } \mathbf{x} \in \mathbb{R}^k \qquad (3.66)$$

where

$$\{\mathbf{X}_{\check{k}} \le \mathbf{x}_{\check{k}}\} = \{X_1 \le x_1, \dots, X_{\check{k}} \le x_{\check{k}}\},$$
$$\{\mathbf{V} \le \mathbf{v}\} = \{V_1 \le v_1, V_2 \le v_2, V_3 \le x_3\},$$
$$\{\mathbf{Z} \le \mathbf{z}\} = \{Z_1 \le z_1, \dots, Z_{\check{k}} \le z_{\check{k}}\}.$$

While the portions relating to the development stage, $\prod_{\tilde{j}=1}^{\tilde{k}} \mathbb{P}\left(U_{\tilde{j}} \le u_{\tilde{j}}\right)$ and $\mathbb{P}\left(\mathbf{X}_{\check{k}} \le \mathbf{x}_{\check{k}}\right)$, can be modified, the portions of the usage stage, $\mathbb{P}\left(\mathbf{V} \le \mathbf{v}\right)$ and $\mathbb{P}\left(\mathbf{Z} \le \mathbf{z}\right)$, are determined and cannot be changed in practice.

The feature space $\mathbb{X}^k$ possesses all combinations of controllable and uncontrollable factors of both the development and usage stage. Due to passively controllable manufacturing tolerances $\mathbb{X}^k$ must be possibly investigated under multiple feature space distributions. Hence, whenever multiple specification tables are considered for manufacturing tolerances - for example, when taking into account two sensor types with different offset error specification limits - both resulting feature space distributions have to be covered in a representative manner by experiments. This requirement limits the applicability of MoBEO in a production and lifetime-oriented process, as the simulation time effort necessary quickly exceeds all time targets. In industry, a well accepted approach is replacing rather slow computer simulation models, like MoBEO, by much faster simulating empirical models or so called metamodels. These models are capable of simulating accurately several thousands of experiments within seconds, if they have been fit in an appropriate manner. Hence, metamodeling approaches require at first the appropriate allocation of experiments in the feature space and their simulation by MoBEO. For this purpose, the generation of such appropriate experiments and their simulation will take the center stage in the next two chapters.

Chapter 4

Design of Experiments

The experimental feature space $\mathbb{X}^k \subset \mathbb{R}^k$, spanned by the factors of the development and usage stage, might be subject to multiple features distributions F^k. Setting the course for a representative coverage of these distributions by design of experiments stands in the main focus of the chapter. For the most part corresponding to Piffl and Stadlober [2015], this chapter will give a guideline on how to efficiently provide high dimensional computer experiments.

Computers have become an indispensable tool in all technical fields where complex processes need to be investigated and optimized. Although computer simulation models, like MoBEO, are becoming more and more precise, the associated increase in complexity has hesitated much faster computation times and has caused confusing frameworks over the last years (cf. Currin et al. [1991] & Siebertz, Van Bebber and Hochkirchen [2010]). DOE minimizes here the simulation effort for investigating the relationship between input factors and response variables $Y_1, \ldots, Y_s$, and it provides moreover the experiments necessary to set up functional relationships by so called metamodels.

In section 4.1 3-level screening designs are compared and the power of the *Definitive Screening design* of Jones and Nachtsheim [2011] is emphasized in the context of computer simulation models. Then, the ***T****otal **S***ensitivity **I***ndex* (TSI) of Homma and Saltelli [1996] is recalled. The TSI allows identifying significant factors for the considered response(s). In section 4.2 the ***L****atin **H***ypercube design* (LH-design) is discussed as well as its capability to provide representative experiments concerning F^k. With regard to the results obtained, our new space-filling design approach, the *Depth-design*, is introduced and presented in section 4.3.

4.1 Statistical Treatment of Computer Experiments

Let $\mathbb{X}^k$ denote the feature space, whose spanning factors have been preselected with regard to response variables $Y_1, \ldots, Y_r$. Then, the random vector $\mathbf{X} = (X_1, \ldots, X_k) \sim F^k$ represents all possible combinations of input factors $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{X}^k$ for MoBEO within $\mathbb{X}^k$. In practice, the simulation capacity N is limited when using MoBEO. This should be expressed by assuming the number of feasible simulations within $25 \leq N \leq 500$ for this chapter. The number N should be optimally selected to build metamodels with regard to $Y_1, \ldots, Y_r$. As illustrated by figure 4.1, this is accomplished by conducting a

screening and a space-filling procedure. It is assumed that most processes or systems are
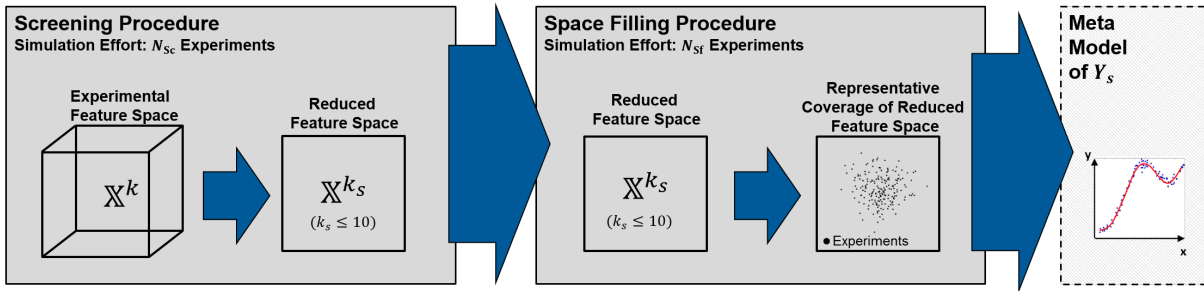
Figure 4.1: Guideline to Generate Experiments for Metamodels

dominated by a few factors. This well-accepted rule of thumb is known in statistics as the "*sparsity-of-effects principle*". Indeed, this rule turns out to be true in most cases (cf. Wu and Hamada [2009], Mee [2009] and Montgomery [2013]). Hence, a preliminary screening procedure, using a portion N_{Sc} of N feasible experiments, can cheaply reduce all considered factors $X_1, \ldots, X_k$ to the most significant factors $X_1, \ldots, X_{k_s}$ (usually $k_s \leq 10$) for each considered response variable Y_s. As a consequence, it is possible to reduce $\mathbb{X}^k$ to a subspace $\mathbb{X}^{k_s}$ for $s = 1, \ldots, r$. The associated reduction of k to k_s eases the experimental coverage by N_{Sf} experiments during the subsequent space-filling procedure.

4.1.1   Screening Designs

Screening designs enable the identification of the most relevant factors $X_1, \ldots, X_k$ of a response variable Y_s with comparable low simulation effort. Given a fixed number of factors k to investigate, screening designs mainly differ in the number of experiments N_{Sc} and in the interpretability of the estimated effects (cf. Montgomery [2013]). Both criteria are important, because, beside the numerical effort, the estimated effects outline how single factors or combinations of factors influence a response variable Y_s. Dependent on the design approach, the estimated effects may bias each other. With regard to screening designs, the statistician is especially interested in *clear* main effects. Such clear main effects are neither biased by other main effects nor by 2nd order effects (cf. Montgomery [2013]). While 2-level screening designs, share the idea of comparing the results of two different factor levels "−1" and "+1" (extreme cases), 3-level screening designs additionally consider the factors at a center level "0". This makes it possible to estimate pure quadratic effects that may indicate a possible curvature in the relationship between input factors and response variable.

During the screening procedure, factors $X_1, \ldots, X_k$ need to be assumed as independent random variables. This simplification may be especially feasible for computer simulation models. Due to the neglected correlation structure, in the worst case, more factors may remain after the screening procedure. Hence, although ambient temperature and ambient humidity, for instance, strongly correlate in a positive manner, they are assumed as independent for the screening procedure. If the independent setting of two or more factors is not possible, it is suggested to successively neglect such factors until an independent consideration becomes possible. It is proposed to standardize the edges of $\mathbb{X}^k$ to $[-1, +1]$

so that eventually $\mathbb{X}^k = [-1, +1]^k$. This standardization measure is advantageous for two reasons. On the one hand, quantitative input factors can be easily compared. On the other hand, DOE notation of 3-level screening designs uses "-1", "0" and "+1" to denote the factor levels. Design matrices of a 2^3 and 3^2 designs are given for illustration purposes by table 4.1 and 4.2. Scientists and engineers do often feel more comfortable with 3-
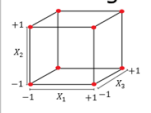
2^3- Design	X_1	X_2	X_3	y_n
Experiment 1	-1	-1	-1	y_1
Experiment 2	+1	-1	-1	y_2
Experiment 3	-1	+1	-1	y_3
Experiment 4	+1	+1	-1	y_4
Experiment 5	-1	-1	+1	y_5
Experiment 6	+1	-1	+1	y_6
Experiment 7	-1	+1	+1	y_7
Experiment 8	+1	+1	+1	y_8

Table 4.1: 2^3 Design
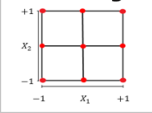
3^2- Design	X_1	X_2	y_n
Experiment 1	-1	-1	y_1
Experiment 2	0	-1	y_2
Experiment 3	+1	-1	y_3
Experiment 4	0	-1	y_4
Experiment 5	0	0	y_5
Experiment 6	0	+1	y_6
Experiment 7	+1	-1	y_7
Experiment 8	+1	0	y_8
Experiment 9	+1	+1	y_9

Table 4.2: 3^2 Design

level screening designs, because they tend to expect a substantial non-linear relationship between input factors and the response. As a result, the straightforward application of 3-level full factorial designs (3^k designs) is not possible, because the number of experiments N_{Sc} exponentially grows in k. For instance, the $3^{k=10}$ design requires already $N_{\text{Sc}} = 59049$ experiments. The following alternative 3-level screening designs need less experiments for k factors.

1. 3^{k-p} fractional factorial design [Montgomery, 2013]

2. Box-Behnken design [Box and Behnken, 1960]

3. Face-Centered Central Composite design [Box and Wilson, 1951]

4. Definite Screening design [Jones and Nachtsheim, 2011]

At closer inspection, the 3^{k-p} fractional factorial design comes along with a major disadvantage, as each effect has $(3^p - 1)/2$ different aliases. If the number of fractions p is increased in order to decrease the number of experiments, the generation of 3^{k-p} fractional factorial designs, exhibiting clear main effects, becomes very challenging. Xu [2005] proposes a method, based on coding theory in order to identify clear main effects of large 3^{k-p} fractional factorial designs. However, his method is only applicable for $k \leq 20$ factors. The very efficient Box-Behnken design, does not provide the vertices of $\mathbb{X}^k$. This property may be especially useful for applications where factor combinations with extreme levels are not realistic. Still, if the factors are allowed to follow other than normal distributions, this property is not desirable at all.

An often proposed alternative to a 3^k design is the 3-level Face-Centered Central Composite design. In order to cope with non-linearity issues, it expands a 2-level full factorial design (2^k design) by introducing axial experiments and multiple experiments in the center of the feature space. Nevertheless, the 2^k design yields too many experiments when k is large. One may have the idea to replace the 2^k design by a 2^{k-p} fractional factorial design with clear main effects. This measure, however yields significantly biased pure quadratic effects, which contradicts the original intended purpose of a 3-level screening design (cf. Siebertz, Van Bebber and Hochkirchen [2010]).

Cheng and WU [2001], Tsai, Gilmour and Mead [2000] and Jones and Nachtsheim [2011] address their studies to *non-regular*[1] 3-level designs for screening procedures. The work of Jones and Nachtsheim and their **D**efinitive **S**creening design (DS-design) with $N_{Sc} = 2 \cdot k + 1$ experiments should be emphasized at this point. Based on the strategy to minimize the correlation structure among second order effects by the *coordinate exchange algorithm*[2] and the target to achieve uncorrelated main effects by the *fold-over technique*[3], they produced 3-level designs with orthogonal columns. Jones and Nachtsheim itemize their characteristics in the following way:

1. The number of required experiments is only one more than twice the number of considered factors ($N_{Sc} = 2 \cdot k + 1$).

2. Unlike resolution III designs, main effects are completely independent of two-factor interactions.

3. Unlike resolution IV designs, two-factor interactions are not completely confounded with other two-factor interactions, although they may be correlated.

4. Unlike resolution III, IV and V designs with added center experiments, all quadratic effects are estimable in models comprised of any number of linear and quadratic main-effects terms.

5. Quadratic effects are orthogonal to main effects and not completely confounded (though correlated) with interaction effects.

Table 4.3 shows the orthogonal design matrix of a DS-design for 4 input factors. There, the included fold-over technique becomes clearly visible, as each even numbered row is obtained by multiplying each value of the previous row by −1. The last experiment denotes always the center experiment. If the DS-design of table 4.3 is projected into three dimensions, one may get an idea which experiments of the feature space are simulated. The projected DS-design is depicted by figure 4.2. However, Jones and Nachtsheim [2011] constructed their designs with a complicated algorithm, which was not able to find *orthogonal designs*[4] for $k = 12$. Furthermore, this algorithm suffered from slow convergence speed in cases where k was large. In the paper of Xiao, Lin and Bai [2012], the generation of the DS-design was completed by a well applicable construction method based on conference matrices. Although conference matrices do not exist for k factors, if $k \equiv 2 \mod 4$

---

[1]cf. Montgomery [2013]
[2]cf. Goos and Jones [2011]
[3]cf. Wu and Hamada [2009]
[4]cf. Montgomery [2013]

DS-Design 4 Factors	X_1	X_2	X_3	X_4	y_n
Experiment 1	0	+1	+1	+1	y_1
Experiment 2	0	-1	-1	-1	y_2
Experiment 3	+1	0	+1	-1	y_3
Experiment 4	-1	0	-1	+1	y_4
Experiment 5	+1	-1	0	+1	y_5
Experiment 6	-1	+1	0	-1	y_6
Experiment 7	+1	+1	-1	0	y_7
Experiment 8	-1	-1	+1	0	y_8
Experiment 9	0	0	0	0	y_9

Table 4.3: DS-Design for 4 Factors
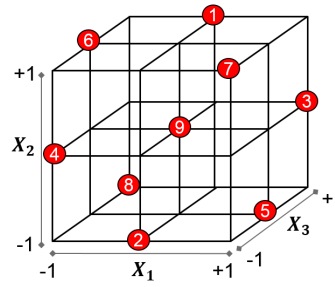
Figure 4.2: DS-Design for 4 Factors, Projected to 3 Dimensions

and $k - 1 \neq a^2 + b^2$ for a and b integers (cf. Van Lint and Seidel [1966]), it is always possible to increase k, construct a DS-design and finally delete the dispensable columns. As long as no design rows are deleted, orthogonality is not affected by this procedure. Altogether, these designs provide the possibility to estimate pure main effects, quadratic effects and interaction effects with one run-through. Hence, the DS-design is a very efficient screening tool for all kinds of processes, where a non-linear relationship between many input factors and one response is expected.

4.1.2 Evaluation of the Definitive Screening Design

By construction, the non-regular DS-design provides $2 \cdot k + 1$ degrees of freedom. In principle, this allows the estimation of $2 \cdot k$ different effects. Still, the assessment of all main effects, 2nd order interactions and quadratic effects requires $2 \cdot k + \binom{k}{2} + 1$ degrees of freedom. Here, pair-by-pair comparisons are a possible remedy to estimate all effects (cf. Wu and Hamada [2009]). For two different factors $A \coloneqq X_i$ and $B \coloneqq X_j$ ($1 \le i < j \le k$) the following linear ("L") and quadratic ("Q") effects can be estimated.

$$\underbrace{A_L, A_Q, B_L, B_Q, AB_{L \times L},}_{\text{of interest}} \overbrace{AB_{L \times Q}, AB_{Q \times L}, AB_{Q \times Q}}^{\text{not considered}}. \tag{4.1}$$

While the pure linear, pure quadratic and simple interaction effects are of interest, the effects $AB_{L \times Q}$, $AB_{Q \times L}$, $AB_{Q \times Q}$ are neglected for the screening procedure. These effects are usually negligible, hard to interpret and should be rather used to build regression models. Thus, the square sums of the *contrasts*[5], which can be directly derived by the estimating $A_L, A_Q, B_L, B_Q, AB_{L \times L}$, are compared to the total sum of squares of Y_s. Consequently, one compares the impact of single effects on the variability of the response. All effects of interest can be estimated, if this so called **ANalysis Of VAriance** (ANOVA) is conducted sequentially for all $\binom{k}{2} = \frac{k \cdot (k-1)}{2}$ factor pairs.

Let $y_1, \ldots, y_{N_{\text{Sc}}}$ denote the results of response variable Y_s, obtained by the DS-design.

---

[5]cf. Montgomery [2013]

Then, the linear contrast C_{A_L} of factor A is computed by the contrast weights of Hinkelman and Kempthorne [1994] as follows.

$$C_{A_L} = -1 \cdot \sum_{n=1}^{N_{Sc}} y_n \cdot \mathbf{1}_{\{A=-1\}} + 0 \cdot \sum_{n=1}^{N_{Sc}} y_n \cdot \mathbf{1}_{\{A=0\}} + 1 \cdot \sum_{n=1}^{N_{Sc}} y_n \cdot \mathbf{1}_{\{A=+1\}} = -1 \cdot y_- + 0 \cdot y_0 + 1 \cdot y_+. \quad (4.2)$$

The wrapped fold-over technique of the DS-design yields clear main effects. The square sum of a linear effect is obtained by the fraction of its squared contrast and the number of experiments in each group, weighted by its squared contrast coefficient (cf. Montgomery [2013]). For A_L that means

$$SS_{A_L} = \frac{C_{A_L}^2}{\sum\limits_{n=1}^{N_{Sc}} \mathbf{1}_{\{A=-1\}} \cdot (-1)^2 + \sum\limits_{n=1}^{N_{Sc}} \mathbf{1}_{\{A=+1\}} \cdot (+1)^2} = \frac{C_{A_L}^2}{|y_-| \cdot (-1)^2 + |y_+| \cdot (+1)^2}. \quad (4.3)$$

Estimating the pure quadratic effect of a factor requires experiments at the center level "0". As the DS-design only provides three center level experiments per factor, but $(k-1)$ experiments with levels "-1" and "$+1$", the design is generally not balanced. By definition of a contrast, the scalar product between vector m, which contains the observations per level

$$m = (|y_-|, |y_0|, |y_+|), \quad (4.4)$$

and the contrast weight vector $c = (c_{-1}, c_0, c_{+1})$ must be zero (cf. Montgomery [2013]). Hence, to comply with $m^T c \overset{!}{=} 0$, it is necessary to standardize the elements of c to the smallest coefficient of m, which corresponds to $|y_0| = 3$ for the DS-design. Hence, the contrast weight vector for the quadratic contrast $c = (1, -2, 1)$ needs to be replaced by $c = \left(1 \cdot \frac{|y_0|}{|y_-|}, -2 \cdot \frac{|y_0|}{|y_0|}, 1 \cdot \frac{|y_0|}{|y_+|}\right)$ so that

$$m^T c = (k-1) \cdot \frac{3}{(k-1)} - 2 \cdot 3 + (k-1) \cdot \frac{3}{(k-1)} = 0 \quad (4.5)$$

holds for the DS-design. The quadratic contrast C_{A_Q} of factor A can be estimated by

$$[C_{A_Q}] = c_{-1} \cdot y_- + c_0 \cdot y_0 + c_1 \cdot y_+ = 1 \cdot \frac{3}{(k-1)} \cdot y_- + (-2) \cdot y_0 + 1 \cdot \frac{3}{(k-1)} \cdot y_+. \quad (4.6)$$

Although orthogonal contrasts are used, it has to be pointed out that the contrast C_{A_Q} is biased due to the correlation structure of the DS-design. The contrasts describe the coherence between the biased $[C_{A_Q}]$ and C_{A_Q} as follows

$$[C_{A_Q}] = C_{A_Q} + \sum_{E \neq A} r_{(A_Q, E_Q)} \cdot C_{E_Q} + \sum_{F \neq E} r_{(A_Q, EF_{L \times L})} \cdot C_{EF_{L \times L}}, \quad (4.7)$$

where factors E and F differ from factors A and B. Jones and Nachtsheim [2011] showed that the absolute value of correlation $r_{(\cdot, \cdot)}$ between pure quadratic effects converges towards $\frac{1}{3}$ and between quadratic effects and 2nd order interaction effects towards $\frac{1}{6}$ for $k \to \infty$.

The square sum of the pure quadratic effects is estimated by

$$[SS_{A_Q}] = \frac{([C_{A_Q}])^2}{|y_{-1}| \cdot \left(\frac{3}{(k-1)}\right) \cdot (+1)^2 + |y_0| \cdot (-2)^2 + |y_{+1}| \cdot \left(\frac{3}{(k-1)}\right) \cdot (+1)^2}. \quad (4.8)$$

Finally, the contrasts of the two-factor interaction effects have to be determined. Here, the contrast weights origin a (3×3)-matrix, derived from the matrix multiplication of the contrast weights of the involved effects

$$c_{AB_{L \times L}} = c(A_L)^T \times c(B_L) = (-1, 0, 1)^T \times (-1, 0, 1) = \begin{pmatrix} +1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & +1 \end{pmatrix}. \qquad (4.9)$$

As a result, the contrast $C_{AB_{L \times L}}$ of the two-factor interaction is estimated by

$$[C_{AB_{L \times L}}] = 1 \cdot y_{--} + (-1) \cdot y_{-+} + (-1) \cdot y_{+-} + 1 \cdot y_{++} \qquad (4.10)$$

whereas

$$[C_{AB_{L \times L}}] = C_{AB_{L \times L}} + \sum_E r_{(AB_{L \times L}, E_Q)} \cdot C_{E_Q} + \sum_{\substack{E \neq F \\ EF \neq AB}} r_{(AB_{L \times L}, EF_{L \times L})} \cdot C_{EF_{L \times L}}. \qquad (4.11)$$

The desired square sum is

$$[SS_{AB_{L \times L}}] = \frac{([C_{AB_{L \times L}}])^2}{|y_{--}| \cdot (+1)^2 + |y_{-+}| \cdot (-1)^2 + |y_{+-}| \cdot (-1)^2 + |y_{++}| \cdot (+1)^2}. \qquad (4.12)$$

4.1.3 Total Sensitivity Index

The total sum of squares $SST = \sum_{n=1}^{N_{Sc}} (y_n - \bar{y})^2$ reflects the artificially generated variability of response Y_s for a number of conducted experiments. Whenever two factor pairs, (A, B) and (A, C) are evaluated by the described pair-by-pair approach, SST is partitioned into the following components

$$SST = SS_{A_L} + [SS_{A_Q}]_B + SS_{B_L} + [SS_{B_Q}]_A + [SS_{AB_{L \times L}}] + SSR_{AB} \text{ and} \qquad (4.13)$$
$$SST = SS_{A_L} + [SS_{A_Q}]_C + SS_{C_L} + [SS_{C_Q}]_A + [SS_{AC_{L \times L}}] + SSR_{AC}, \qquad (4.14)$$

where the error sums of squares SSR_{AB}, SSR_{AC} reflect the portion of SST, not explained by the effects considered. The combination of factor A with another factor $E \neq A$ entails $k - 1$ different square sums $[SS_{A_Q}]_E$. Here, it is proposed to take their mean value

$$\overline{[SS_{A_Q}]} = \frac{1}{k - 1} \cdot \sum_{E \neq A} [SS_{A_Q}]_E \qquad (4.15)$$

as the desired statistic. With the TSI, introduced by Homma and Saltelli [1996], the portion of the total variability caused by factor A can be finally expressed by

$$ST_A = \frac{SS_{A_L} + \overline{[SS_{A_Q}]} + \sum_{E \neq A} [SS_{AE_{L \times L}}]}{SST}. \qquad (4.16)$$

Originally, this index demands unbiased estimators of the involved square sums. In the worst case $\overline{[SS_{A_Q}]}$ or $[SS_{AE_{L\times L}}]$ are affected by an overall noise, which may obscure relevant effects or erroneously increase the significance of unimportant effects (cf. figure 4.3). A possible remedy is to consider a bias corrected TSI, where the overall noise is removed with respect to a relevance level $0 < \alpha \leq 0.1$.

$$ST_{A_{\mathrm{corr}(\alpha)}} = \frac{SS_{A_L} + \overline{[SS_{A_Q}]}_{\mathrm{corr}(\alpha)} + \sum\limits_{E\neq A} [SS_{AE_{L\times L}}]_{\mathrm{corr}(\alpha)}}{SST} \qquad (4.17)$$

with

$$\overline{[SS_{A_Q}]}_{\mathrm{corr}(\alpha)} := \begin{cases} \delta_{A_Q} & \alpha \leq \dfrac{\overline{[SS_{A_Q}]}}{SST} \\ 0 & \text{else} \end{cases}, \qquad (4.18)$$

where

$$\delta_{A_Q} = \left| \overline{[SS_{A_Q}]} - \underset{E\neq A}{\mathrm{mean}}\left\{ \overline{[SS_{E_Q}]}\cdot \mathbb{I}_{\left\{ \frac{\overline{[SS_{E_Q}]}}{SST} < \alpha \right\}} \right\} \right|. \qquad (4.19)$$

Accordingly, the same formulas apply to $[SS_{AB_{L\times L}}]_{\mathrm{corr}(\alpha)}$. The impact of this correction becomes visible when figures 4.3 and 4.4 are compared. Figure 4.3 contains fractions $\overline{[SS_{A_Q}]}/SST$ and $[SS_{AB_{L\times L}}]/SST$ that origin a simulated DS-design with $k = 69$ factors. As visible, the TSIs of many factors are likely to be influenced by a multiplicity of interaction effects. On basis of the "*sparsity-of-effects principle*", it is expected that only a few decisive main effects and second order effects dominate the simulated process. Assuming that the correlations of equation (4.11) are the root cause of this problem, the corrected TSI shall be applied. Figure 4.4 illustrates the corresponding fractions after having ap-
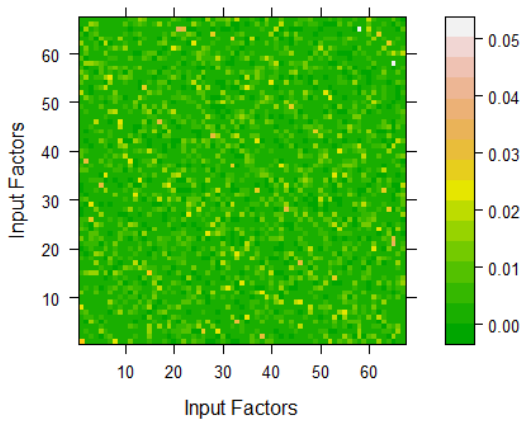
Figure 4.3: Share of Effect Square Sums in the Total Square Sum
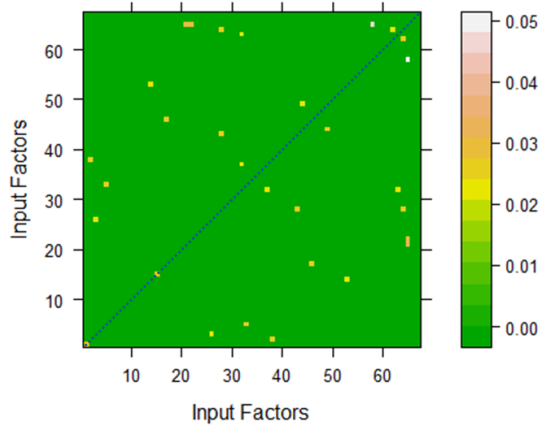
Figure 4.4:   Share of Corrected Effect Square Sums in the Total Square Sum

plied the corrected TSI with $\alpha = 0.025$. As visible, two quadratic effects and only a few two-factor interaction effects are relevant. After a consultation with the model experts these effects have been confirmed. Moreover, no further effects than the identified effects

have been expected to be significant. The results of the concerning corrected TSIs are visible in figure 4.5 and 4.6. Pie plots or bar diagrams are suggested to highlight which factors are significant and which factors can be neglected respectively. As expected, only
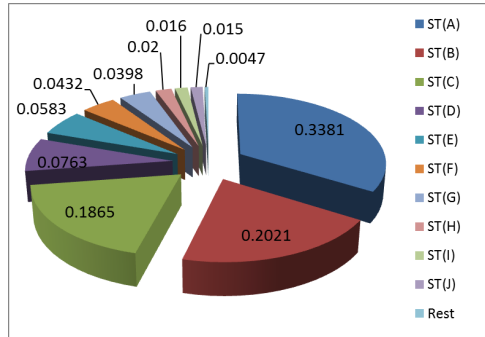
Figure 4.5: Pie Plot - Corrected TSIs
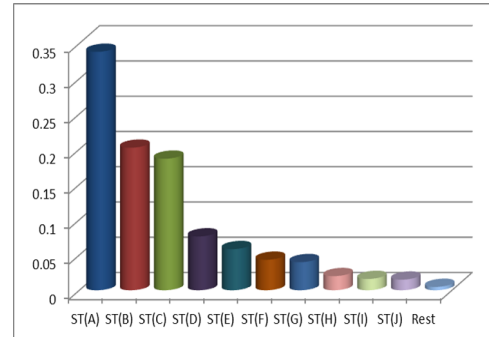
Figure 4.6: Bar Plot - Corrected TSIs

$k_s = 10$ of the $k = 69$ considered factors dominate the MoBEO model in terms of Y_s. To sum up, the DS-design of Jones and Nachtsheim [2011] with $N_{\mathrm{Sc}} = 2 \cdot k + 1$ experiments is a powerful screening tool for computer simulation models, where a non-linear relationship between factors and response variables is expected.

4.1.4 Space-Filling designs

The DS-design of Jones and Nachtsheim [2011] is proposed as the first step of building metamodels for response variables $Y_1, \ldots, Y_r$. Thereby, the simulation of one DS-design allows detecting the most significant factors $X_1, \ldots, X_{k_s}$ and reducing $\mathbb{X}^k$ to subspaces $\mathbb{X}^{k_s}$ for each Y_s, $s = 1, \ldots, r$. Although the space-filling procedure needs to be repeated for each response variable in theory, in practice this only denotes the worst case scenario. Often, two or more response variables share almost the same significant input factors so that the individual reduction of $\mathbb{X}^k$ is not useful. From methodical point of view, it is sufficient to consider one response variable Y_s.

While the feature space distribution F^k played a minor role during the screening procedure, it plays an important role in the subsequent space-filling procedure. Input factors regarding the usage stage are associated to distributions that cannot be modified in practice. Following that, a representative coverage of the reduced feature space $\mathbb{X}^{k_s}$ (in terms of F^{k_s}) should be ensured for all usage stage related factors as applicable. This measure later ensures that the corresponding metamodels work precisely under different usages. Contrary to that, the portions of the development stage of F^{k_s} in (3.66) have to remain variable. The distributions of the factors that model manufacturing tolerances might be known or can be estimated. Still, investigating more than the existent distributions for these factors, requires a precise metamodel for the entire domain. This also applies to the open system parameters, which are modeled by uniform random variables. Due to the fact that the factors of development stage are assumed to be independent to the factors

of the usage stage, it suffices to merge two separate space-filling designs. The one has to provide most efficiently distributed experiments with regard to the development stage. The other must provide most representative experiments concerning the usage stage. On this account well known space-filling design approaches should be briefly discussed.

The first space-filling designs were introduced to uniformly distribute experiments within a given feature space. Hence, the experiments of these designs represent approximately a multidimensional uniform distribution. As Montgomery [2013] points out, this is desirable, if the experimenter is not aware of the relationship between the input factors and the response, and if the underlying feature space distribution is either unknown or not determined. The most popular space-filling design approaches, included in commercial software packages, are listed below in table 4.4. The *pdf* of the k_s-dimensional

Design Type	Background	Non-uniform Distributions possible	Source
1. Sphere Packing	$\max\limits_{\mathbf{x},\mathbf{x}'\in\mathbb{X}^k}\min\|\mathbf{x}-\mathbf{x}'\|_2$	No	Johnson et al. [1990]
2. Uniform	$\min\limits_{\mathbb{X}^k}$ Discrepancy	No	Fang [1980]
3. Maximum Entropy	$\max\limits_{\mathbb{X}^k}$ Shannon Entropy	Yes	Shewry and Wynn [1987]
4. IMSE[6]	$\min\limits_{\mathbb{X}^k}$ MSE(Residuals)	Yes	Sacks et al. [1989]
5. (Quasi) MC	Law of Large Numbers	Yes	Johnson [1987]
6. Latin Hypercube	Permutation Matrix	Yes	McKay et al. [1979]

Table 4.4: Standard Space-Filling Designs

distribution F^{k_s} is obtained by

$$f^{k_s}\left(x_1,\ldots,x_{k_s}\right) = \int\limits_{-\infty}^{\infty} \ldots \int\limits_{-\infty}^{\infty} f^k\left(x_1,\ldots,x_{k_s}\right)\, dx_1' \ldots dx_{k-k_s+1}', \qquad (4.20)$$

where $x_1',\ldots,x_{k-k_s+1}'$ correspond to the insignificant factors $X_1',\ldots,X_{k-k_s+1}'$, eliminated by the screening procedure. Independent of their capability to represent other than uniform distributions, space-filling designs 1., 2., 3. and 4. of table 4.4 are not further discussed in this thesis. This is due to their long computation times in high dimensional spaces.

Given that the factors $X_1,\ldots,X_{k_s}$ are independent random variables (i.e. $X_j \overset{\text{ind}}{\sim} F_j$ for $j = 1,\ldots,k_s$), Fishman [1996] motivates pseudo MC methods in order to represent multidimensional distributions. So called *pseudo MC-designs* can be generated by directly merging random samples of the marginal distributions F_j (cf. Siebertz, Van Bebber and Hochkirchen [2010]). If the assumption of independent factors is too restrictive, there

---

[6]***I****ntegrated* ***M****ean* ***S****quared* ***E****rror*

exists an easy transformation based on the *Cholesky decomposition* to generate representative experiments for a multivariate normal distribution $F^{k_s} \equiv N_{k_s}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (cf. Dagpunar [1988]). Moreover, Johnson [1987] presents several techniques on how to generate random samples of other than multivariate normal distributions, which can be used to generate pseudo MC-designs. Probably because of their simple applicability, the pseudo MC-design has gained a considerable amount of recognition in engineering. Still, it has to be mentioned that the efficiency of such designs is heavily influenced by the assumed distribution. Replicate experiments may especially cause redundant information in regions of the feature space with small variability. Beyond all that, an efficient space-filling design "per se" can not be subject to total randomness. If F^{k_s} is a multidimensional uniform distribution, a more efficient experimental coverage can be achieved by quasi MC-designs (cf. Siebertz, Van Bebber and Hochkirchen [2010]). Because the theory of non-uniform random numbers does not directly apply to the quasi MC-design, this technique is not suggested, however, for other than uniform distributions.

The LH-design of McKay et al. [1979] is also a well accepted space-filling design approach in industry. The general idea behind the LH-design is to cover the feature space as best as possible, while avoiding replicate experiments at the same time. Although the LH-design is the first proposed space-filling design, and it was originally intended, like other pioneer space-filling designs, to generate experiments in a uniform manner, Stein [1987] introduced an approach to generate LH-designs for general multidimensional distributions. If factors $X_1, \ldots, X_{k_s}$ are independent, the LH-design is generated by partitioning $\mathbb{X}^{k_s}$ into hyper-rectangles with equal probability. Eventually, the experiments are distributed over the resulting lattice so that the projection to any one dimension yields exactly one experiment per cell. For cases, where factors $X_1, \ldots, X_{k_s}$ are not independent, Stein [1987] proposes a technique using inverse marginal *cdfs* and permutation matrices to generate appropriate LH-designs. Figure 4.7 illustrates the lattice approach for two independent factors $X_1 \sim N\left(0, \frac{1}{4}\right)$ and $X_2 \sim N\left(0, \frac{1}{4}\right)$. Figure 4.8 depicts the LH-design, generated by the approach of Stein [1987], for two dependent factors (X_1, X_2), subject to the bivariate normal distribution $N^2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = c\,(0,0) \ \text{ and } \ \boldsymbol{\Sigma} = \begin{pmatrix} \frac{1}{4} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{4} \end{pmatrix}. \tag{4.21}$$

The LH-design approach is a simple method on how to involve the feature space distribution and avoid replicates at the same time. Still, Stein [1987] advises large sample sizes in order to achieve representativeness for the underlying feature space distribution.

To sum up, the LH-design of McKay et al. [1979] is the perfect tool to efficiently generate experiments regarding the factors of the development stage. The question remains whether or not the LH-design of Stein [1987] can produce representative experiments with regard to a multidimensional distribution, when the simulation capacity is limited. In the next section this question shall be answered by the aid of multidimensional quantiles. For simplicity, it is assumed that all remaining factors $X_1, \ldots, X_{k_s}$ after the screening procedure relate to the usage stage.
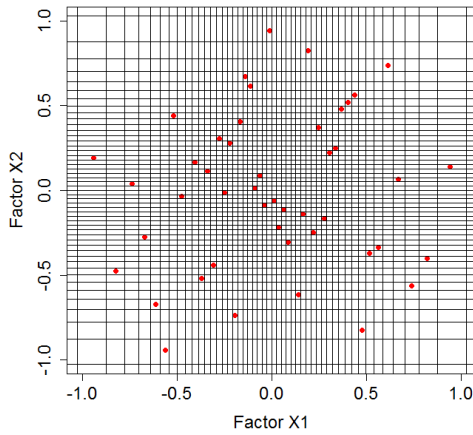
Figure 4.7: LH-Design for Independent Factors
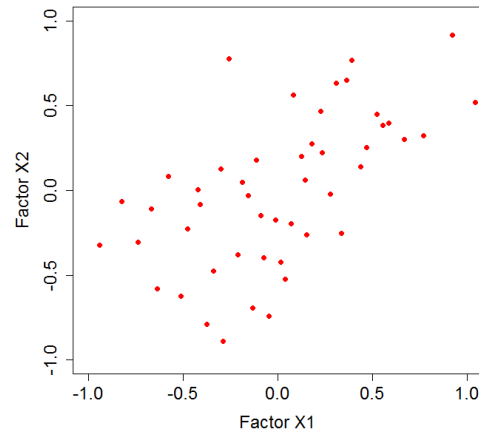
Figure 4.8: LH-Design for Dependent Factors

4.2 Multidimensional Quantiles and Depth Functions

In one dimension the representativeness of a set of points with regard to a particular distribution can usually be illustrated by a **Quantile-Quantile** (QQ) plot, which scatters the empirical quantiles against the theoretical quantiles. However, the concept of one-dimensional quantiles with their natural linear order is not directly transferable to multi-dimensional distributions. Serfling [2010] overcomes this issue by using median-oriented quantiles. These imply nested contours, which enclose regions containing amounts of inner probability around the k_s-dimensional median of the multidimensional distribution F^{k_s}. Median-oriented quantiles are defined for $\mathcal{F}$, that is the class of distributions on the *Borel sets* of $\mathbb{R}^{k_s}$.

Definition 4.1. (*Median-Oriented Quantile Function*)
For $\boldsymbol{\vartheta} \in \mathbb{B}^{k_s-1}(\mathbf{0}) = \{\boldsymbol{\vartheta} \in \mathbb{R}^{k_s} : \|\boldsymbol{\vartheta}\| \le 1\}$ the median-oriented quantile function $Q(\cdot,\cdot) : \mathbb{B}^{k_s-1}(\mathbf{0}) \times \mathcal{F} \to \mathbb{R}^{k_s}$ generates nested contours $\{\mathbf{x} = Q(\boldsymbol{\vartheta}, F^{k_s}) : \|\boldsymbol{\vartheta}\| = c, 0 \le c \le 1\}$ in $\mathbb{R}^{k_s}$ that enclose the k_s-dimensional median, defined by $\mathbf{M}_{F^{k_s}} = Q(\mathbf{0}, F^{k_s})$.

The following example shows how to determine the median-oriented quantiles for a bivariate normal distribution, modeling two independent random variables (trivial covariance matrix $\boldsymbol{\Sigma}$).

Example 4.1. (*Median-Oriented Quantiles of a Bivariate Normal Distribution*)

Let $N^2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote a bivariate normal distribution, well defined by the parameters

$$\boldsymbol{\mu} = (0,0) \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 = \sigma^2 & 0 \\ 0 & \sigma_2^2 = \sigma^2 \end{pmatrix} \text{ for } 0 < \sigma < \infty. \tag{4.22}$$

Then, the circles A_l, defined by radii r_l and by the center point $\boldsymbol{\mu} \in \mathbb{R}^{k_s}$, are contours possessing inner probabilities $0 < p_l < 1$. These are given by

$$\mathbb{P}(\mathbf{X} \in A_l) = p_l \text{ for } l \in \mathbb{R}, \tag{4.23}$$

where $p_l < p_{l'}$, if $l < l'$ (or $A_l \subset A_{l'}$). For $l \in \mathbb{R}$ and $\mathbf{x} \in \overline{A_l}$ the transformation

$$R\left(\mathbf{x}, N^2\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)\right) = p_l \cdot \frac{\mathbf{x}}{\|\mathbf{x}\|} = \boldsymbol{\vartheta} \tag{4.24}$$

yields the mapping $R : \mathbb{R}^{k_s} \times \mathcal{F} \to \mathbb{B}^{k_s-1} \times \mathcal{F}$, which satisfies $\|\boldsymbol{\vartheta}\| = p_l < 1$. The probability p_l can be determined for a circle A_l with radius r_l as follows.

$$p_l = \int_{A_l} f^2_{N^2(\mathbf{0}, \boldsymbol{\Sigma})}(x_1, x_2)\, d(x_1, x_2) = \frac{1}{2\pi\sigma^2} \cdot \int_{A_l} \exp\left(-\frac{1}{2\sigma^2} \cdot (x_1^2 + x_2^2)\right) d(x_1, x_2).$$

Substitution of $x_1 = r \cdot \cos\phi$ and $x_2 = r \cdot \sin\phi$ by polar coordinates (r, ϕ) yields

$$p_l = \frac{1}{2\pi\sigma^2} \cdot \int_{\phi=0}^{2\pi} \int_0^{r_l} r \cdot \exp\left(-\frac{1}{2\sigma^2} \cdot (r^2 \cdot \cos^2\phi + r^2 \cdot \sin^2\phi)\right) dr\, d\phi.$$

Noting that $\cos^2\phi + \sin^2\phi = 1$, one receives the Circular Gaussian probabilities

$$p_l = 1 - \exp\left(-\frac{r_l^2}{2\sigma^2}\right).$$

The inverse of transformation (4.24) yields the median-oriented quantile function, with which the median-oriented quantiles are obtained for all inner probabilities $0 < p_l < 1$. They are given by

$$\left\{ \mathbf{x} = \frac{\boldsymbol{\vartheta} \cdot \overbrace{\|\mathbf{x}\|}^{=r_l}}{p_l} : \sqrt{\vartheta_1^2 + \vartheta_2^2} = p_l \right\} \tag{4.25}$$

for $\boldsymbol{\vartheta} \in \mathbb{B}^{k_s-1}(\mathbf{0})$ and $p_l \neq 0$. The k_s-dimensional median $\mathbf{M}_{F^{k_s}} = Q(\mathbf{0}, F^{k_s})$ corresponds to $\boldsymbol{\mu}$.

The solution approach shown above on how to determine the median-oriented quantiles of a distribution F^{k_s} is a special application of the **D**epth-**O**utlyingness-**Q**uantile-**R**ank (D-O-Q-R) *paradigm*, described in detail by Serfling [2010]. In a nutshell, it says that three functions, the *depth function*, the *outlyingness function* and the *centered rank function*, closely relate to the median-oriented quantile function. Serfling [2010] uses this relation in order to show that it suffices to find a depth function $D(\mathbf{x}, F^{k_s})$ to obtain a valid quantile function $Q(\boldsymbol{\vartheta}, F^{k_s})$ for a continuous distribution F^{k_s}. Thereby, the depth function must possess nested contours, enclosing a point $\mathbf{M}_{F^{k_s}} \in \mathbb{R}^{k_s}$ and bounding regions

$$A_l = \left\{\mathbf{x} \in \mathbb{R}^{k_s} : D(\mathbf{x}, F^{k_s}) \geq l\right\} (l \in \mathbb{R}). \tag{4.26}$$

As already indicated by example 4.1, where the nested contours A_l were implicitly assumed in (4.23), the median-oriented quantile function $Q(\boldsymbol{\vartheta}, F^{k_s})$, $\boldsymbol{\vartheta} \in \mathbb{B}^{k_s-1}$, has an inverse at each $\mathbf{x} \in \mathbb{R}^{k_s}$, given by the centered rank function $R(\mathbf{x}, F^{k_s})$. The magnitude $O(\mathbf{x}, F^{k_s}) = |R(\mathbf{x}, F^{k_s})|$ of the centered rank function is called outlyingness function. The outlyingness function, in turn, is related with the depth function by $D(\mathbf{x}, F^{k_s}) = 1 - O(\mathbf{x}, F^{k_s})$.

After Serfling [2010] a valid quantile function is induced by setting $\mathbf{M}_{F^{k_s}} = Q(\mathbf{0}, F^{k_s})$ and $\boldsymbol{\vartheta} = R(\mathbf{x}, F^{k_s}) = p \cdot \mathbf{v} \in \mathbb{B}^{k_s-1}$, where $\mathbf{v}$ is the unit vector that points towards $\mathbf{x}$ from $\mathbf{M}_{F^{k_s}}$. Moreover, p_l denotes the inner probability of region A_l with $\mathbf{x}$ on its boundary (cf. transformation (4.24)). Hence,

$$\mathbb{P}\left(\mathbf{X} \in A_l\right) = p_l \text{ for } \mathbf{X} \sim F^{k_s}. \tag{4.27}$$

Furthermore, the probability p_l, $l \in \mathbb{R}$, equals to the outlyingness of all $\mathbf{x}$ on the closure $\overline{A_l}$ of A_l, i.e. $O(\mathbf{x}, F^{k_s}) = \|\boldsymbol{\vartheta}\| = p_l$. Consequently, $\overline{A_l}$ corresponds to the median-oriented quantile possessing an inner probability p_l. Zuo and Serfling [2000] define the depth function in a more general way.

Definition 4.2. (*Depth Function*)
Let the mapping

$$D(\cdot, \cdot) : \mathbb{R}^{k_s} \times \mathcal{F} \to \mathbb{R}_+ \tag{4.28}$$

be bounded, non-negative and satisfying the subsequent conditions.

(i) *$D(\mathbf{Ax} + \mathbf{b}, F^{k_s}_{\mathbf{AX}+\mathbf{b}}) = D(\mathbf{x}, F^{k_s}_{\mathbf{X}})$ holds for any random vector $\mathbf{X} \sim F^{k_s}$, any non-singular matrix $\mathbf{A} \in \mathbb{R}^{k_s \times k_s}$ and any vector $\mathbf{b} \in \mathbb{R}^{k_s}$ (affine invariance property);*

(ii) *$D(\boldsymbol{\theta}, F^{k_s}) = \sup_{\mathbf{x} \in \mathbb{R}^{k_s}} D(\mathbf{x}, F^{k_s})$ holds for any $F^{k_s} \in \mathcal{F}$ having a "center" $\boldsymbol{\theta}$ in $\mathbb{R}^{k_s}$ (maximality at center);*

(iii) *For any $F^{k_s} \in \mathcal{F}$ having a "center" $\boldsymbol{\theta} \in \mathbb{R}^{k_s}$, $D(\mathbf{x}, F^{k_s}) \le D(\boldsymbol{\theta} + \delta(\mathbf{x} - \boldsymbol{\theta}), F^{k_s})$ holds for all $\delta \in [0, 1]$ (monotonicity relative to the deepest point);*

(iv) *$D(\mathbf{x}, F^{k_s}) \to 0$ as $\|\mathbf{x}\| \to \infty$ for arbitrary $F^{k_s} \in \mathcal{F}$ (vanishing at infinity).*

Then, $D(\cdot, F^{k_s})$ is called statistical depth function.

A random vector $\mathbf{x} \in \mathbb{R}^{k_s}$ is usually denoted as centrally symmetric about a "center" $\boldsymbol{\theta}$, if $\mathbf{x} - \boldsymbol{\theta} \overset{d}{=} \boldsymbol{\theta} - \mathbf{x}$, whereas "$\overset{d}{=}$" means equal in distribution. Zuo and Serfling [2000] introduce a more general notion, which defines $\mathbf{X}$ to be halfspace symmetric about $\boldsymbol{\theta}$, if $\mathbb{P}(\mathbf{X} \in H) \ge \frac{1}{2}$ for every closed halfspace $H \subset \mathbb{R}^{k_s}$ that contains $\boldsymbol{\theta}$.

With condition (ii) of definition 4.2, the D-O-Q-R paradigm implies that the center $\boldsymbol{\theta}$ of F^{k_s} equals to the multidimensional median, i.e. $\boldsymbol{\theta} \equiv \mathbf{M}_{F^{k_s}}$. Thus, it follows that the k_s-dimensional median possesses maximal depth.

With moderate effort it is possible to determine the probabilities p_l in (4.27) of the k_s-dimensional normal distribution $N^{k_s}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with parameters

$$\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{k_s}) \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_{k_s}^2 \end{pmatrix} \text{ for } |\boldsymbol{\Sigma}| < \infty. \tag{4.29}$$

As highlighted by Waugh [1961], the determination procedure requires the integration of the *pdf* of F^{k_s} over a centered k_s-dimensional ellipse, whose k_s semi axes correspond to the diagonal elements of $\boldsymbol{\Sigma}$. The analytical resolvability of the median-oriented quantile

task, as presented in example 4.1, becomes challenging for $F^{k_s} = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($\boldsymbol{\Sigma}$ not trivial), because the orientation of the k_s-dimensional ellipse changes. Moreover, unless a multi-variate normal distribution is considered, the analytical resolvability of probability p_l of region A_l is no longer ensured.

The D-O-Q-R paradigm and the depth function approach shall be used to estimate median-oriented quantiles of a general distribution F^{k_s} by realizations $\mathbf{x}_1, \ldots, \mathbf{x}_{N'}$ of $\mathbf{X}_n \overset{\text{iid}}{\sim} F^{k_s}$. This is because, the depth function $D(\mathbf{x}_n, F^{k_s})$ yields a center-outward ordering of $\mathbf{x}_1, \ldots, \mathbf{x}_{N'}$. The probability p_l of the region A_l, as defined in (4.26), is estimated as follows.

$$\mathbb{P}(\mathbf{X} \in A_l) = p_l \approx \frac{|A_l|}{N'} \text{ for } l \in \mathbb{R} \text{ and } \mathbf{X} \sim F^{k_s}. \tag{4.30}$$

The median oriented quantiles, which correspond to probabilities p_l, are estimated by the convex hulls of A_l. In the papers of Liu et al. [1999] and Zuo and Serfling [2000] various depth functions are presented, compared and discussed in terms of the four desired properties *affine invariance, maximality at center, monotonicity relative to the deepest point* and *vanishing at infinity*. Zuo and Serfling [2000] conclude by recommending the halfspace- and the projection depth, because they fulfill the four desired properties of definition 4.2. Although the D-O-Q-R paradigm can be applied for every valid depth function, an appropriate choice would be the affine invariant version of the L_2 depth. This recommendation is threefold:

1. The affine invariant version of the L_2 depth fulfills all properties of definition 4.2.

2. The affine invariant version of the L_2 depth can be approximated with controllable effort, as shown in (4.33).

3. The affine invariant version of the L_2 depth can be determined for $F^{k_s} \equiv N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Definition 4.3. (L_2 *Depth*)
The L_2 depth of $\mathbf{x} \in \mathbb{R}^{k_s}$ is given by

$$L_2 D(\mathbf{x}, F^{k_s}) = \left(1 + \mathbb{E}_{F^{k_s}} \|\mathbf{x} - \mathbf{X}\|_2\right)^{-1}, \tag{4.31}$$

where $\mathbf{X} \sim F^{k_s}$.

Mosler [2013] showed that the nested contours, possessed by (4.31), bound convex regions A_l. In addition to this desirable characteristic, Zuo and Serfling [2000] describe that the L_2 depth provides all desired properties with the exception of affine invariance. Here, Rao [1988] have demonstrated that an affine invariant version of definition 4.3 can be obtained by replacing the euclidean norm with $\|\mathbf{x}\|_M \equiv \sqrt{\mathbf{x}^T \mathbf{M} \mathbf{x}}$, where $\mathbf{M} \in \mathbb{R}^{k_s \times k_s}$ is a positive definite matrix.

Definition 4.4. (*Affine Invariant Version of the L_2 Depth*)

$$\tilde{L}_2 D(\mathbf{x}, F^{k_s}) = \left(1 + \mathbb{E}_{F^{k_s}} \|\mathbf{x} - \mathbf{X}\|_{\boldsymbol{\Sigma}^{-1}}\right)^{-1}, \tag{4.32}$$

where $\mathbf{X} \sim F^{k_s}$ and $\boldsymbol{\Sigma}^{-1}$ the covariance matrix of distribution F^{k_s}.

The expected norm in definition 4.4 can be estimated by a moderate sized subset $\{\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_{N_{L_2}}\}$ of $\{\mathbf{x}_1, \ldots, \mathbf{x}_{N'}\}$ in $N' \cdot N_{L_2} < N'^2$ operations as follows.

$$\tilde{L}_2 D(\mathbf{x}_n, F^{k_s}) \approx \left(1 + \frac{1}{N_{L_2}} \cdot \sum_{n=1}^{N_{L_2}} \sqrt{(\mathbf{x} - \tilde{\mathbf{x}}_n)^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \tilde{\mathbf{x}}_n)} \right)^{-1} \text{ for } n = 1, \ldots, N'. \quad (4.33)$$

If $F^{k_s} \equiv N^{k_s}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the expected value in (4.32) can be determined. For $\mathbf{x}_n \in \mathbb{R}^{k_s}$, the exact solution requires the consideration of the positive random variable

$$W_{\mathbf{x}_n} = (\mathbf{x}_n - \mathbf{X})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - \mathbf{X}) \in \mathbb{R}_+, \text{ where } \mathbf{X} \sim N^{k_s}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \quad (4.34)$$

Given that the distribution $F_{W_{\mathbf{x}_n}}$ of $W_{\mathbf{x}_n}$ is known, we can apply the well known formula

$$\mathbb{E}(W_{\mathbf{x}_n}^\gamma) = \gamma \cdot \int_0^\infty t^{\gamma-1} \cdot \mathbb{P}(W_{\mathbf{x}_n} > t) \, dt \quad (4.35)$$

for $\gamma = \frac{1}{2}$. Now it holds that distribution $F_{W_{\mathbf{x}_n}}$ corresponds to a *non-central chi-squared distribution* with k_s degrees of freedom and non-centrality parameter λ^2 (cf. Anderson [1984]).

$$W_{\mathbf{x}_n} \sim \chi^2_{k_s}(\lambda^2) \text{ where } \lambda^2 = (\mathbf{x}_n - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_n - \boldsymbol{\mu}). \quad (4.36)$$

Still, it has to be pointed that the resolvability of the affine invariant version of the L_2 depth must not be confused with the resolvability of the median-oriented quantile task. The product of N' and the computational effort, necessary to evaluate the integral in (4.35), dominates the operations required to estimate the median-oriented quantiles of $F^{k_s} \equiv N^{k_s}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $F^{k_s} \not\equiv N^{k_s}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the computation speed mainly depends on N' and on the approximation quality N_{L_2} in (4.33).

Figure 4.9 illustrates the median-oriented quantiles of a bivariate normal distribution $N^2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with parameters

$$\boldsymbol{\mu} = (\tfrac{1}{2}, \tfrac{1}{2}) \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \frac{1}{8} & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{8} \end{pmatrix} \quad (4.37)$$

that have been estimated by $N' = 10.000$ pseudo MC samples for $p_l = 0.1, \ldots, 0.9$. Representatively for a distribution $F^{k_s} \not\equiv N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, figure 4.10 shows the nonparametric estimation approach for the versatile Beta distribution. The estimation was based on $N' = 10.000$ pseudo MC samples $(x_1, x_2) \in \mathbb{R}^2$, which were realizations of

$$\frac{X_1}{X_1 + X_2 + X_3} \text{ and } \frac{X_2}{X_1 + X_2 + X_3}, \quad (4.38)$$

where

$$X_1 \sim \gamma(2, 5) \text{ and } X_2 \sim \gamma\left(\tfrac{1}{2}, 5\right) \text{ and } X_3 \sim \gamma(2, 5). \quad (4.39)$$

$\gamma(\kappa, \theta)$ denotes a gamma distribution with shape parameter $\kappa > 0$ and scale parameter $\theta > 0$. The computation of the convex hulls was performed with the algorithm of Barber et al. [1996].

It should be clarified which sample size N' is required to estimate the median-oriented quantiles of a distribution F^{k_s} in an adequate manner. By means of the law of large
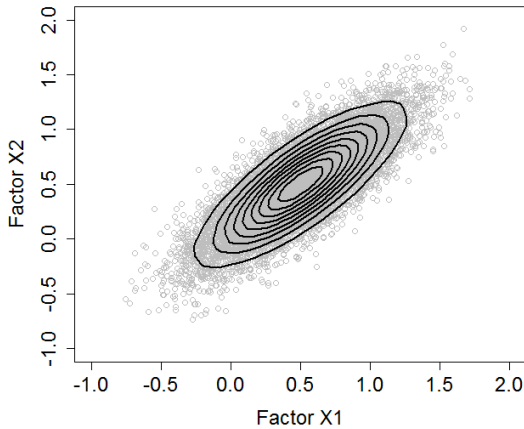
Figure 4.9: Median-Oriented Quantiles of $N^2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
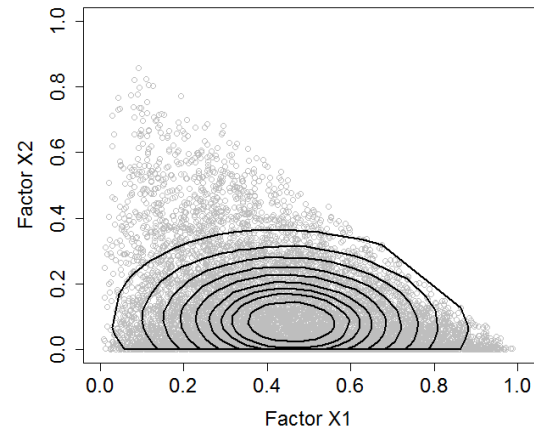
Figure 4.10: Median-Oriented Quantiles of Bivariate Beta Distribution

numbers, one can count on stochastic sampling strategies like the LH-design (or pseudo MC sampling) to depict the true distribution, if N' is large enough. The estimated median-oriented quantiles can be easily validated with a test set $\{\mathbf{x_1}, \ldots, \mathbf{x}_{N_\text{test}}\}$ $(N_\text{test} > N')$, because the estimated probabilities p_l of all regions A_l can be compared to the fractions p_l^test of the test set A_l^test, given by

$$A_l^\text{test} = \left\{\mathbf{x} \in \{\mathbf{x_1}, \ldots, \mathbf{x}_{N_\text{test}}\} : D(\mathbf{x}, F^{k_s}) \geq l\right\}, \text{ and } p_l^\text{test} = \frac{|A_l^\text{test}|}{N_\text{test}}. \tag{4.40}$$

The performance of the estimated median-oriented quantiles shall be assessed by the **R**oot **M**ean **S**quared **E**rror (RMSE)

$$RMSE_s = \sqrt{\frac{1}{N'} \cdot \sum_l \left(p_l - p_l^\text{test}\right)^2}. \tag{4.41}$$

Indeed, $RMSE_s$ is a measure of representativeness of the realizations $\mathbf{x}_1, \ldots, \mathbf{x}_{N'}$ for the distribution F^{k_s}. If the Latin-Hypercube sampling approach of Stein [1987] is used to generate the concerning realizations the open question, stated at the end of subsection 4.1.4, can be answered. That question was whether or not the LH-design is descriptive enough to represent a distribution F^{k_s}, when the number of experiments is limited.

Example 4.2. (*Validation of Estimated Median-Oriented Quantiles: LH-Design*)

The validation procedure should be demonstrated for the example of the multivariate normal distribution $F^{k_s} \equiv N^{k_s}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. This is because random normal distributions are easily generated and the affine invariant version of the L_2 depth can be analytically determined. Several experiments with different normal distributions $N^{k_s}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ showed that $N' = 150.000$ is a good choice for $k_s \leq 100$ dimensions. As an example, figure 4.11 illustrates the validation of median-oriented quantiles of $N^{100}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ estimated by $N' = 150.000$ experiments and validated by $N_{test} = 250.000$ experiments.
Hence, a LH-design with $N' = 150.000$ experiments represents $N^{100}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ in an adequate

manner. Even if $N' = 150.000$ are far beyond the considered boundary of $N_{Sf} = N - N_{Sc}$ experiments, $\mathbf{x}_1, \ldots, \mathbf{x}_{150.000}$ can be definitely used to validate median-oriented quantiles, estimated by less experiments. On this account,

$$N' \in \left\{ \underbrace{25, 50, 75, 100, 150, 200, 300, 400, 500,}_{possible\ values\ for\ N_{Sf}} \underbrace{1.000, 10.000, 50.000, 100.000}_{for\ validation\ purpose} \right\}$$

and $N_{test} = 150.000$ are considered. As outlined in subsection 4.1, the dimension of the initial feature space $\mathbb{X}^k$ will practically reduce to $k_s \leq 10$ after the screening procedure. Figure 4.12 highlights the performance of the median-oriented quantiles, when estimated on basis of a LH-design, comprising $N' \in \{25, 50, 100, 150, 500\}$ experiments. It is clearly visible that the larger k_s is given, the more fails the LH-design to represent the underlying distribution F^{k_s} for a given N'. This is due to the construction method of Stein [1987], which "only" considers marginal cdfs. Figure 4.13 illustrates that for the multivariate normal distribution, considered with different dimensions $k_s \in \{2, 5, 10, 20, 50\}$. For each dimension, the outlined $RMSE_s$ are based on an average of 20 different distributions. The LH-design fails to accurately estimate the median-oriented quantiles, when $k_s > 10$ and the $N_{Sf} \leq 500$.
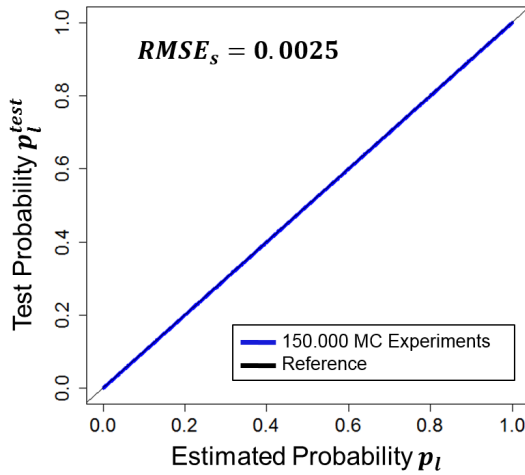
Figure 4.11: Validation of Estimated Median-Oriented Quantiles of $N^{100}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
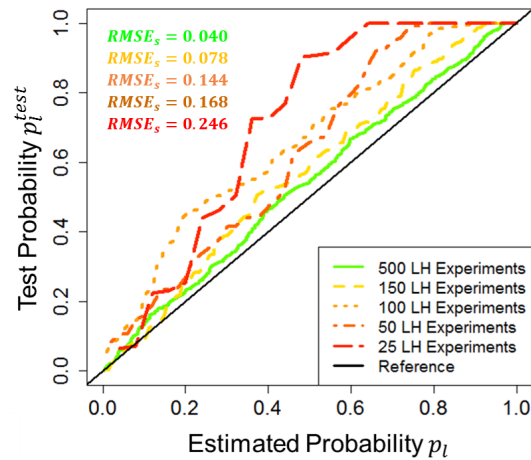
Figure 4.12: Validation of Estimated Median-Oriented Quantiles of $N^{10}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

The LH-design of Stein [1987] is a very efficient design that includes the feature space distribution. Still, it lags in representing a high dimensional distribution, when less than 500 experiments are feasible. Hence, the LH-design of Stein [1987] is not the best choice to generate representative experiments for the factors of the usage stage. On this account, a new space-filling design, expanding the LH-design approach of Stein [1987] by median-oriented quantiles and depth functions, is now developed. This so called *Depth-design* will be introduced in the subsequent section.
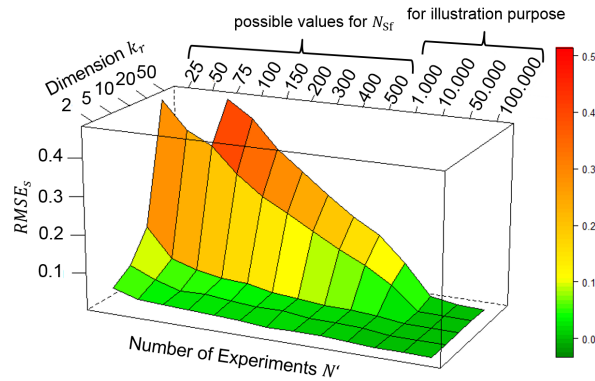
Figure 4.13: LH-Design: Validation of Estimated Median-Oriented Quantiles of $N^{k_s}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

4.3 The Depth-Design

The combination of a high dimensional feature space and a limited number of experiments significantly endangers the capability of the LH-design of Stein [1987] to represent an underlying feature space distribution. In such cases ($N_{\mathrm{Sf}} \leq 500$ and $k_s \approx 10$) it is assumed that the experiments, generated by the LH-design, do not provide the information necessary to represent the associated distribution F^{k_s} in an accurate manner. This disadvantage can be eliminated by our new Depth-design, which expands the LH-design of Stein [1987] by the median-oriented quantile approach of Serfling [2010].
In the first step of constructing the Depth-design, realizations $\mathbf{x}_1, \ldots, \mathbf{x}_{N'}$ of $\mathbf{X} \sim F^{k_s}$ need to be generated by a stochastic sampling procedure (i.e. pseudo MC sampling, LH-design, etc.) in order to estimate the median-oriented quantiles of the underlying distribution. In doing so, it is proposed to choose N' in correspondence with figure 4.13 so that $RMSE_s$ is kept at a minimum level. This means, one has to set $N' = 50.000$, if $k_s = 50$ for instance. If N_{Sf} experiments are feasible, the median-oriented quantiles $\overline{A}_l$ are of particular interest. They correspond to the probabilities

$$p_l = \frac{\overline{A}_l}{N'} \in \left\{ \frac{0}{N_{\mathrm{Sf}} - 1}, \frac{1}{N_{\mathrm{Sf}} - 1}, \ldots, \frac{N_{\mathrm{Sf}} - 1}{N_{\mathrm{Sf}} - 1} \right\}. \tag{4.42}$$

The regions A_l are received by sorting $\mathbf{x}_1, \ldots, \mathbf{x}_{N'}$ in an ascending order by their depth value. Then, the estimated median-oriented quantiles $\overline{A}_l$ can be approximated by

$$\overline{A}_l \approx \left\{ \mathbf{x} \in \{\mathbf{x}_1, \ldots, \mathbf{x}_{N'}\} : D(\mathbf{x}, F^{k_s}) - \epsilon \leq l \leq D(\mathbf{x}, F^{k_s}) + \epsilon \right\} (\epsilon > 0). \tag{4.43}$$

The idea is to adjust the LH-design of Stein [1987], given by N_{Sf} experiments $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_{N_{\mathrm{Sf}}}$ by the estimated median-oriented quantiles, obtained by condition (4.42). By construction, each experiment of the LH-design is linked to a specific $\overline{A}_l$. Starting with the smallest probability p_l and successively proceeding to the largest probability, the experiments of the Depth-design are generated as follows.

$$DD(\mathbf{x}_n) = \underset{\mathbf{x} \in \overline{A}_l}{\arg\min} \|\mathbf{x} - \tilde{\mathbf{x}}_n\|_2 \text{ for } n = 1, \ldots, N_{\mathrm{Sf}}. \tag{4.44}$$

The computational effort of the Depth-design is dominated by the estimation procedure of the median-oriented quantiles, as described in section 4.2.

Example 4.3. (*Generation of the Depth-Design for Dimension $k_s = 2$*)

The reader may reconsider the bivariate normal distribution in (4.37) and the bivariate Beta distribution respectively in (4.38). There, the affine invariant version of the L_2 depths were estimated from $N' = 10.000$ pseudo MC samples. The introduced Depth-design approach is illustrated by figures 4.14 and 4.15. Due to visualization purposes, the number of experiments corresponds to $N_{Sf} = 10$ only. Both figures clearly highlight the idea of the design by the yellow arrows. When figure 4.12 is compared with figure 4.16,
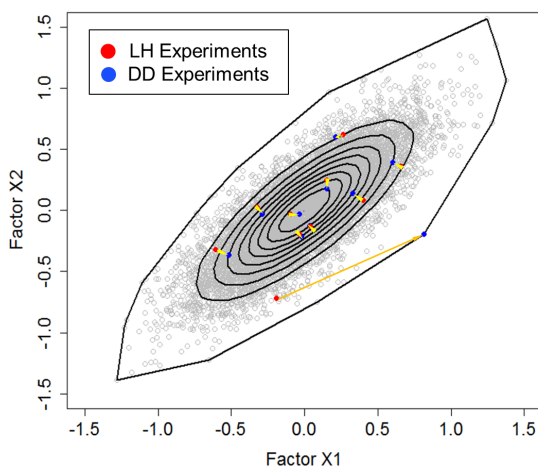
Figure 4.14: Depth-Design for Bivariate Normal Distribution $N^2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
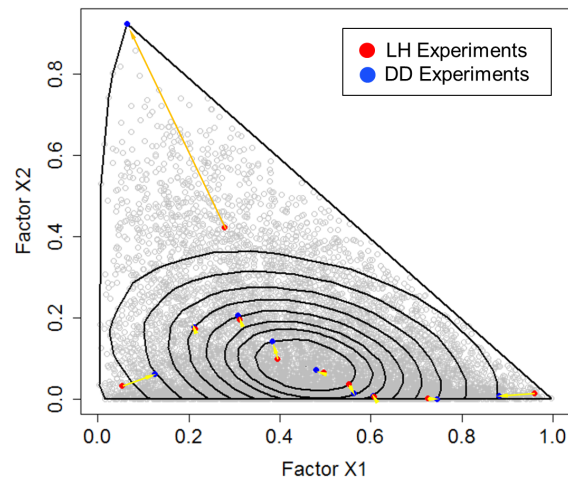
Figure 4.15: Depth-Design for Bivariate Beta Distribution

the improvement achieved by the Depth-design becomes instantly visible. According to the $RMSE_s$ obtained, already 25 Depth-design experiments contain more information about the considered distribution ($k_s = 10$) as 500 experiments, generated by the LH-design of Stein [1987]. Again, the median-oriented quantiles have been validated with a test set of size $N_{test} = 150.000$.

    In line with figure 4.13, the performance of the Depth-design shall be analyzed, when the dimension k_s is increased and the number of experiments is varied. Figure 4.17 clearly shows the power of the Depth-design in terms of representing a multidimensional distribution with a limited number of experiments. Individual tests showed that the performance for non-normal distributions is comparable to the results obtained in figure 4.17. Additionally, due to the fact that the Depth-design is based on the LH-design, it nearly shares all desirable advantages, like good projection properties or the efficient distribution of the experiments within $\mathbb{X}^{k_s}$. Compared to the LH-design, the generation of a Depth-design is often more time-consuming. Nevertheless, the capability of the Depth-design to represent an underlying distribution with much less experiments can save hundreds to thousands of experiments. For these reasons the application of the Depth-design is
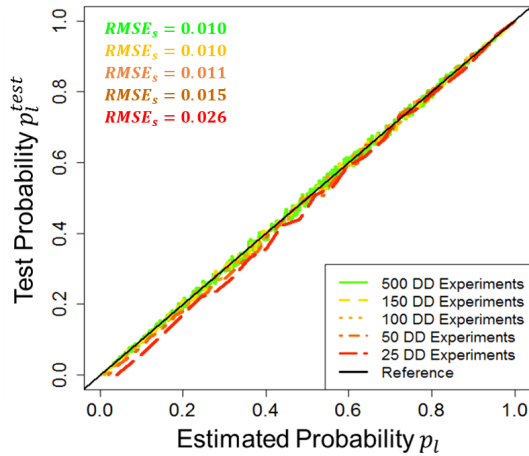
Figure 4.16: Depth-Design for Bivariate Normal Distribution $N^2 (\boldsymbol{\mu}, \boldsymbol{\Sigma})$
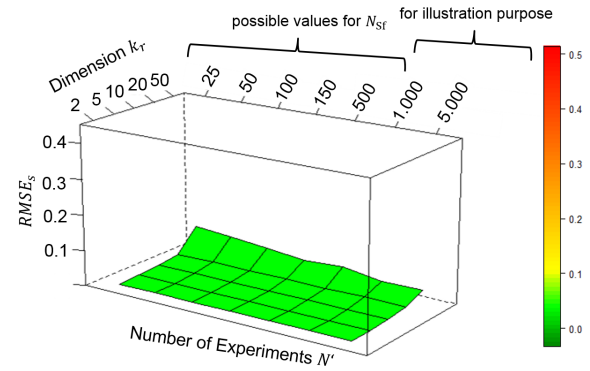
Figure 4.17: LH-Design: Validation of Estimated Median-Oriented Quantiles of $N^{k_s} (\boldsymbol{\mu}, \boldsymbol{\Sigma})$

proposed for all factors of the usage stage during the space filling procedure.

In this chapter, a guideline was provided on how to efficiently generate experiments for computer simulation models, when the number of feasible experiments is limited ($N \leq 500$). It was assumed that the simulation models have a large number ($k \geq 50$) of input factors, subject to a multidimensional distribution F^k. As the first step of generating representative experiments it was suggested to spend a portion N_{Sc} of N for a screening procedure. Here, the very efficient Definitive Screening design of Jones and Nachtsheim [2011] was proposed. This was because pure linear and 2nd order effects were assessable with only $N_{Sc} = 2 \cdot k + 1$ experiments. Due to the undersized and unbalanced Definitive Screening design, a bias corrected pair-by-pair comparison strategy was established. This strategy enabled the estimation of the corrected TSI of Homma and Saltelli [1996], which gave a relative significance measure for each input factor. The space-filling procedure was demonstrated on the example of one response variable Y_s. It was proposed to merge two independent space-filling designs with $N_{Sf} = N - N_{Sc}$ experiments. While the LH-design of McKay et al. [1979] is perfectly suited for the factors of the development stage, the LH-design approach was not suitable for the factors of the usage stage. As shown for the multivariate normal distribution and the affine invariant version of the L_2 depth, the LH-design of Stein [1987] lagged in transmitting high dimensional distributions to the generated experiments, if $N_{Sf} \leq 500$. Here, the **M**ean **S**quared **E**rror (MSE) between the inner probabilities of the estimated median-oriented quantiles (on basis of $\mathbf{x}_1, \ldots, \mathbf{x}_{N_{Sf}}$) and the inner probabilities of estimated median-oriented quantiles (on basis of pseudo MC samples $\mathbf{x}_1, \ldots, \mathbf{x}_{N'}$, $N' = 150.000$ for $k_s \leq 50$) were consulted as a measure of representativeness with regard to F^{k_s}. A new space-filling design approach, the "Depth-design", was proposed instead to generate the experiments for the factors of the usage stage. It expanded the LH-design of Stein [1987] by the concept of multidimensional quantiles and depth functions. Particularly, the Depth-design adjusted the experiments of the LH-design of Stein [1987] in a way that the median-oriented quantiles were accurately

estimated. Indeed, N_{Sf} = 25 Depth-design experiments provided already more information about a multidimensional distribution than 500 experiments, generated by the LH-design of Stein [1987]. This is because already a few Depth-design experiments carried almost all information about the median-oriented quantiles of the underlying distribution. After having modeled the factors of the development and usage stage in a probabilistic manner in chapter 3, and after having discussed most efficient DOE approaches in order to set up accurate metamodels in this chapter, the next chapter will cope with simulation of the generated experiments by MoBEO.

Chapter 5

Lifetime Simulations

Balancing the TCO of a vehicle series requires the development of a closed mathematical method, able to instantly combine quality related development decisions with their associated risks in the field. This thesis addresses this challenge by the production and lifetime-oriented development process, introduced in chapter 2. Due to the parallel consideration of the development and the usage stage of the product's life cycle, a high dimensional feature space $\mathbb{X}^k$ is spanned in the first phase of this process. Beside the high dimension k of $\mathbb{X}^k$, the underlying distribution F^k shall be varied so that a representative exploration of $\mathbb{X}^k$ requires simulating millions of experiments. Due to the limited simulation capacity of the semi-physical simulation tool MoBEO, accurate metamodels shall be applied in order to overcome the high numerical effort. The experimental basis for the therefore necessary simulation data was given in the previous chapter. This chapter will deal with the simulation of the generated experiments under the consideration of one or multiple system aging states.

Section 5.1 will expand the feature space by the factor "*time*" and will particularly cope with the new state of an engine environment. Section 5.2 will extend the consideration of the factor time to multiple aging levels. Particularly, it will be shown how to embed aging statuses obtained by aging models into the overall simulation process.

5.1 Simulating the New State

Data gathering procedures, like the simulation of experimental designs, are the basis for understanding and modeling unknown systems. Montgomery [2013] cites the great New York Yankees catcher *Yogi Berra* that said "... you can observe a lot just by watching". And in fact, probably almost everyone would sign that a better understanding of an unknown system is easily achieved when comparing system inputs with the corresponding system outputs. The best understanding about an unknown system is achieved, however, when the input factors are not only observed, but systematically controlled. As discussed in chapter 3, the possibility of easily changing input factors makes MoBEO to a very powerful tool in this context. This is not only because modifications of the engine environment are quickly implemented, but also because different operational conditions as well as system aging can be easily simulated.

Let $\mathbb{X}^k$ denote the feature space, yet spanned by the factors of the development and the

usage stage in respect of $Y_1, \ldots, Y_r$. It is of particular interest to investigate the factor combinations of the development and the usage stage over the time axis. This amendment expands $\mathbb{X}^k$ by the additional factor "*time*", as highlighted in figure 5.1. In the course
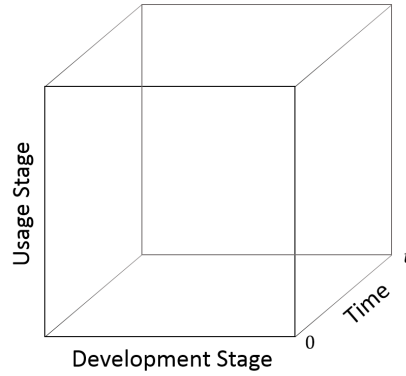
Figure 5.1: Expansion of the Feature Space by the Factor "*time*"

of this section, "*time*" shall be initially assumed as a *dummy* factor, which can only take level "0" within an observation period $[0, t]$. Thereby, it is agreed that the factor level denotes the operation hours of a new engine environment. Hence, if "*time*"= 0, for example, experiments regarding the development and usage stage shall be simulated with a new engine environment. Accordingly to the outlined thoughts in chapter 3, the factor "*time*" can be considered as uniform random variable.

In the course of this section the unknown system to be investigated should correspond to a new engine environment (i.e. "*time*"= 0). The system inputs evolve from random vector (3.64), which can be practically rewritten to

$$\mathbf{X} = (X_1, \ldots, X_k) \sim F^k. \tag{5.1}$$

The rows of a specifically generated experimental design correspond to a systematic series of N realization vectors of (5.1), where the n-th realization $\mathbf{x}_n = (x_{n1}, \ldots, x_{nk})$ is denoted by "experiment n". As pointed out in figure 5.2, the specified design shall now be augmented by an additional "*time*" column in order to highlight the aging state of the simulating engine environment. In this regard, parameter vector $\mathbf{x}^0$ implements the new state "0" of the modeled engine environment. The successive simulation of the design rows
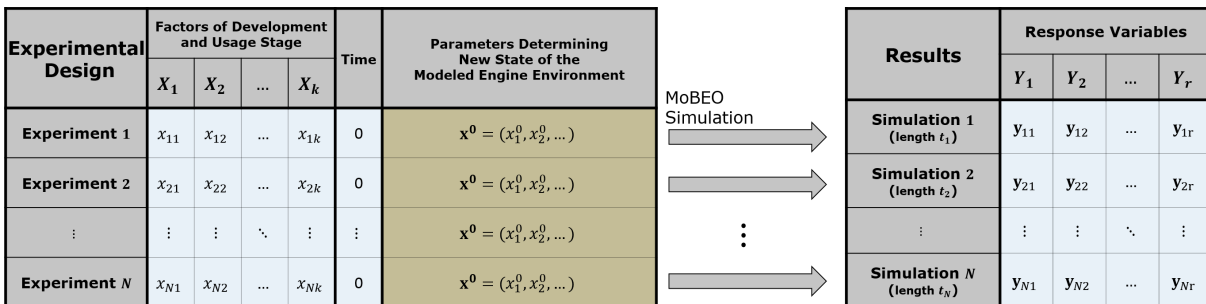
Figure 5.2: Simulating the New State with MoBEO

yields the system output, which consists of the results of response variables $Y_1, \ldots, Y_r$. In particular, "experiment n" yields a set of vectors $\{\mathbf{y}_{n1}, \ldots, \mathbf{y}_{nr}\}$, whose lengths correspond to the lengths of the simulated usage profile $t_{\breve{n}_{\breve{j}}}$ (in s) multiplied by the specified record resolution, outlined in Hertz. Consequently, the simulation of the n-th experiment yields a vector $\mathbf{y}_{ns} \in \mathbb{R}^{t_{\breve{n}_{\breve{j}}} \cdot \mathrm{Hz}}$ for all $s = 1, \ldots, r$, if a usage profile $\breve{j}$ with length $t_{\breve{n}_{\breve{j}}}$ is simulated. In order to ease the notation, the simulated time span of "experiment n" shall be denoted by t_n. The complete notation shall be clarified by the subsequent example.

Example 5.1. (*Simulating with MoBEO*)

It is assumed that two response variables

- $Y_1 \ldots NO_x$ *tailpipe emissions*

- $Y_2 \ldots NH_3$ *tailpipe emissions*

shall be investigated by $N = 4$ MoBEO experiments for an engine environment. One is particularly interested in simulating the air path control strategy with a real air flow sensor and virtual air flow sensor respectively. At the same time, the non-linearity error of the NO_x sensor, ambient conditions and usage behavior shall be altered. Thus, the following input factors are considered.

- $X_1 (= U_1) \ldots$ *Optional Air Flow Sensor*

- $X_2 \ldots$ *Non-Linearity Error NO_x Sensor*

- $X_3 (= V_1) \ldots$ *Ambient Temperature*

- $X_4 (= V_2) \ldots$ *Ambient Humidity*

- $X_5 (= Z) \ldots$ *Usage Profile*

For these input factors a possible experimental design is given by figure 5.3. There, "experiment 1", for instance, corresponds to the objective engine environment without air flow sensor and with a NO_x sensor, exhibiting a completely negative deviated non-linearity error. Simulating "experiment 1" predicts the engine operation with regard to the respective environmental conditions and the usage profile "city driving", which lasts for $t_{\breve{n}_0} = 4h$ for all available city driving observations (i.e. $t_{\breve{n}_0} = 4h$ for all $\breve{n}$). Due to the agreement made above, $t_1 = t_{\breve{n}_0} = 4h$ follows. If the record resolution is set to $10\,Hz$, a vector of length $4 \cdot 3.600 \cdot 10 = 144.000$ is obtained for both response variables. Correspondingly, the lengths of the other results equal to 162.000, to 18.000 and to 144.000 for the other experiments.

Operating a new engine environment for a certain time span causes the underlying system to age in an extent that depends on how the system is operated in the meantime. Beside individual driving habits, system aging does moreover depend on the quality of the materials and on external issues, such as environment conditions.

Experimental Design	Air Flow Sensor	NOx Sensor Non-Linearity Error	Ambient Temperature	Ambient Humidity	Usage Behavior	Time		Results (10 Hz)	Y_1	Y_2
Domain	$\check{\jmath} = 0, 1$: 0 ... No Sensor 1 ... With Sensor	$[-1, +1]$	$[-7.0°C, 37.5°C]$	$[0\,g/kg, 15\,g/kg]$	$\check{\jmath} = 0, 1, 2$: 0 ... City / $t_{\bar{n}_0} = 4.0h\,\forall\bar{n}$ 1 ... Highway / $t_{\bar{n}_1} = 4.5h\,\forall\bar{n}$ 2 ... Legislative WHTC / $t_{\bar{n}_2} = 0.5h\,\forall\bar{n}$	0 ... New State	MoBEO Simulation			
Experiment 1	0	-1	2.50	4.54	0	0		Simulation 1 $t_1 = 4.0h$	$y_{11} \in \mathbb{R}^{144000}$	$y_{12} \in \mathbb{R}^{144000}$
Experiment 2	0	$+1$	19.65	12.09	1	0		Simulation 2 $t_2 = 4.5h$	$y_{21} \in \mathbb{R}^{162000}$	$y_{22} \in \mathbb{R}^{162000}$
Experiment 3	1	-1	-0.85	3.56	2	0		Simulation 3 $t_3 = 0.5h$	$y_{31} \in \mathbb{R}^{18000}$	$y_{32} \in \mathbb{R}^{18000}$
Experiment 4	1	$+1$	11.14	5.57	0	0		Simulation 4 $t_4 = 4.0h$	$y_{41} \in \mathbb{R}^{144000}$	$y_{42} \in \mathbb{R}^{144000}$

Figure 5.3: Data Gathering with MoBEO - Example

5.2   Simulating Multiple Aging States

From worse sensor measurement accuracies to working fluid aging, or from mechanical abrasion of engine components to catalytic performance losses in the exhaust aftertreatment system, material deterioration concerns almost every part of the engine environment. The need for including system aging into a lifetime-oriented development process is self-explanatory, at least for parts of the engine environment that are modeled by MoBEO. So far the factor "*time*" has been only considered for the new state of the modeled engine environment (i.e. "time"= 0). In contrast to the new state, an aged state after t hours (i.e. "time"= t) always depends on how the engine environment has been operated in the meantime. This relationship shall be clarified by the aid of aging models, which transform recordings of damage causing factors into corresponding damages statuses. Subsequently, it will be described on how to arrange the factor "time", when considered on multiple levels, with the recommended screening and space-filling procedure of chapter 4.

5.2.1   Aging Models

The Weibull distribution is a well accepted approach in the industrial sector to model aging behavior (cf. Rinne [2009]). It establishes the theoretical link between the input factors in (5.1) and system aging. The short introduction into the Weibull distribution, follows an approach on how to combine design of experiments with lifetime-oriented considerations. Rinne [2009] defines the *lifetime* of a technical unit as the time span, for which the same exhibits a certain functionality. The technical unit is said to be *dead*, if the functionality reaches a particular damage status. This is for example the case, when a connecting rod of the crankshaft breaks or when catalytic performance losses in the EAS cause emissions to exceed legislative limits. The lifetime L of a technical unit is considered as a continuous *Weibull* distributed random variable (i.e. $L \sim We(a, b, c)$), whose density is defined by

Definition 5.1. (*Density of the Weibull Distribution - $We(a, b, c)$*)

$$f(l \mid a, b, c) = \frac{c}{b} \cdot \left(\frac{l-a}{b}\right)^{c-1} \cdot \exp\left(-\left(\frac{l-a}{b}\right)^c\right) \text{ for } l \geq a.$$

The physical meaning of the Weibull distribution can be probably understood most easily by the *model of the weakest link*, where a technical unit is stripped down to $\iota = 1, \ldots, I$ vitally important subunits. Thereby, the death of a single subunit ι causes the

complete unit to die. The lifetimes of the subunits are assumed to be independently and identically distributed, i.e. $L_\iota \overset{\text{iid}}{\sim} G_L$. Hence, the lifetime of the technical unit is described by the conditional random variable

$$L \mid I \coloneqq \min_{1 \le \iota \le I} \{L_\iota\} \text{ where } F_{L \mid I}(l) = 1 - (1 - G_L(l))^I . \tag{5.2}$$

Because the limiting distribution $F_L = \lim_{I \to \infty} F_{L \mid I}$ is degenerated, in the early 20th century the search after a non-trivial limiting extreme value distribution of F_L started. Eventually, it was Walloddi Weibull (1887-1979), who solved this problem in Weibull [1939a] as well as Weibull [1939b] by a distribution known as the Weibull distribution today.

As described by Rinne [2009], the parameters a, b and c of the Weibull distribution can be estimated by several techniques, like least-squares, maximum likelihood or Bayesian estimation approaches. All estimation approaches require a sample of realized lifetimes $\mathbf{l} = \{l_1, \ldots, l_{\check{N}}\}$ with regard to a particular unit. As illustrated by figure 5.4, the aging behavior of a technical unit, whether discrete or continuous aging is considered, is generally subject to a certain amount of variability. This is mainly because a series of identical appearing technical units is always subject to manufacturing tolerances and different operational and environmental stresses during use. Hence, the parameters of the Weibull distribution establish a link between the lifetime of a technical unit and its development as well as usage stage. Thereby, the location parameter a is called minimum lifetime or threshold. The density moves over the abscissa when parameter a is exclusively changed without changing its shape. The scale parameter b determines the speed of aging, or also the characteristic life, when altered exclusively on its own. The shape parameter c refers to the form and skewness of the density. Figure 5.5 clearly shows the impact on the density, when the distribution parameters are exclusively altered. Furthermore, an outstanding property is highlighted, that is the right skewness of the Weibull distribution, which indicates higher probabilities for shorter rather than longer lifetimes.

AVL has been developing aging models, capable of predicting not only the lifetimes of
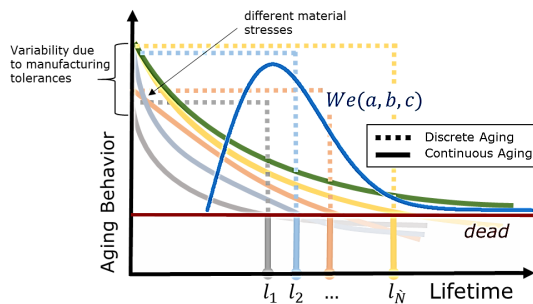
Figure 5.4: Discrete and Continuous Aging Behavior
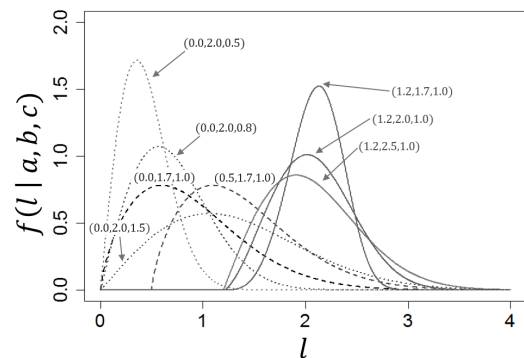
Figure 5.5: Weibull Densities with Different Parameters

individual technical units but also their complete aging behavior. For all that, the Weibull distribution plays a secondary role in this thesis. In general, individual damage modes of

a technical unit are modeled, but not the technical unit itself. Hence, every aging model corresponds to one particular damage mode. The most important aging models of the EAS, for example, are highlighted by table 5.2.1.[1]

Aging / Damage Mode	Modeled EAS Components	Reversibility
Thermal Aging	DOC, TWC, DPF, SDPF, GPF, SCR, ASC	No
Sulfur Poisoning	DOC, TWC, DPF, SDPF, GPF, SCR, ASC	Yes
Chemical Aging	DOC	No
UREA Deposits	upstream SCR	Yes
Ash Poisoning	DPF, SDPF, GPF	No
PGM Migration	SCR	No

Table 5.1: Aging Models of EAS Components

In addition to the EAS, various aging models with respect to important engine components need to be included. Just to mention a few, these are aging models for injector fouling, EGR cooler fouling, turbo charger damages, or divers sensor aging modes.

In general, an aging model $A\left(\cdot \mid \mathbf{p}_A\right)$ transforms a matrix $\boldsymbol{\mathcal{y}}_t^0 \in \mathbb{R}^{t \times \phi}$ into a damage status $D_{\boldsymbol{x}^t}$ at a point in time t (cf. figure 5.6). That is

$$A\left(\boldsymbol{\mathcal{y}}_t^0 \mid \mathbf{p}_A\right) = D_{\boldsymbol{x}^t}, \tag{5.3}$$

where the rows $\left\{\mathbf{y}_1^0, \ldots, \mathbf{y}_\phi^0\right\}$ of $\boldsymbol{\mathcal{y}}_t^0$ denote recordings of damage causing factors of length t, obtained by a new engine environment (outlined here by the superscripted "0" above $\boldsymbol{\mathcal{y}}_t$). The vector $\mathbf{p}_A$ is set by aging model experts. It comprises fit parameters, determining the
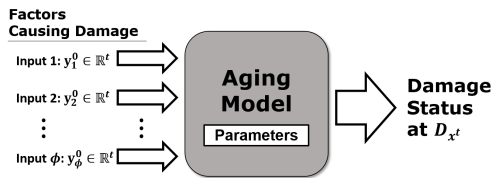
Figure 5.6: General Damage Model
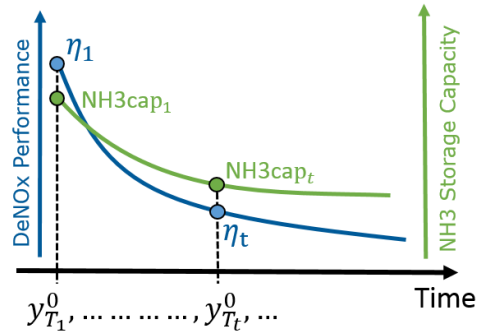
Figure 5.7: Thermal Aging of a SCR Catalyst

prediction performance of the aging model. Given that the modeled engine environment represents the new state, a predicted aged status $D_{\boldsymbol{x}^t}$ is implemented by overwriting $\boldsymbol{x}^0$, set in MoBEO, by parameter vector

$$\boldsymbol{x}^t := \left(x_1^t, \ldots, x_\varphi^t\right) \in \mathbb{R}^\varphi. \tag{5.4}$$

---

[1]cf. appendix for abbreviations

Example 5.2. (*Thermal Aging Model of a SCR Catalyst*)

The irreversible thermal aging of an EAS monolith, for instance, cause coarsening, sintering or phase transformation of the active surface area and its thereon applied catalytic material. Due to the fact that higher temperature loads yield faster catalytic performance decays, the Arrhenius' equation is used for the thermal aging model. The parameter vector $\mathbf{p}_A$ is set in a way that measured performance losses of real catalysts, which had been aged at different temperatures, are well interpolated. The semi-physical nature thereby allows extrapolating measurement data. The resulting parameterized thermal aging model takes a single vector of temperatures as input and converts it to a damage status at a point in time t. If the temperatures are generated by a new engine environment the input vector is denoted by $\mathbf{y}_T^0 = \left(y_{T_1}^0, \ldots, y_{T_t}^0 \right) \in \mathbb{R}^t$. As shown by figure 5.7, the predicted damage status $D_{\boldsymbol{\varkappa}^t}$ of a thermal aging model regarding the SCR catalyst involves among others the $DeNO_x$ performance η_t and the NH_3 storage capacity NH_3cap_t. Consequently, the illustrated damage status in figure 5.7 is implemented by overwriting the modeled engine environment by the parameter vector

$$\boldsymbol{\varkappa}^t = \left(\overbrace{x_1^t}^{\substack{\text{Aged } NH_3 \\ \text{Storage Capacity}}} , \overbrace{x_2^t, x_3^t, x_4^t, \ldots}^{\substack{\text{Aged SCR} \\ \text{Kinetic Parameters}}} , \overbrace{\ldots, x_\varphi}^{\text{Other}} \right) \in \mathbb{R}^\varphi . \tag{5.5}$$

Modeling such irreversible damages, just as thermal aging or PGM migration, provides information about service intervals or warranties. Beside irreversible damages, reversible damages like sulfur poisoning and UREA deposits can be modeled in the same way. Based on adsorption and desorption equations the semi-physical sulfur poisoning model, for instance, transforms a vector of exhaust temperature and lambda values into performance losses of the respective monoliths. The acquired information can be used to adapt counteracting methods, just as regular DeSOx procedures.

Aging models are generally not embedded into the MoBEO environment, meaning that the model environment does not age automatically. This "offline" consideration is mainly argued by the possibility of quickly predicting progressed aging (e.g. $t > 500$h). The therefore necessary recordings $\left\{ \mathbf{y}_1^0, \ldots, \mathbf{y}_\phi^0 \right\}$ are usually replicated to a desired length t. If $\mathbf{y}_\phi^0 \in \mathbb{R}^{t_n}$, for instance, then $\mathbf{y}_\phi^0$ is concatenated $\lceil t/t_n \rceil$ times and cut to a length t.

5.2.2 Implementing Aging Behavior

Let $A_{j'}$, $j' = 1, \ldots, k'$, denote the available parameterized aging models that shall be consulted for faithfully reflecting the usage stage. Then, the damage statuses at a point in time t to be implemented are determined by

$$A_{j'} \left(\boldsymbol{y}_{t,j'}^0 \,|\, \mathbf{p}_{A_j'} \right) = D_{\boldsymbol{\varkappa}_{j'}^t} \text{ for } j' = 1, \ldots, k'. \tag{5.6}$$

For simplicity, it is assumed that damage statuses $D_{\boldsymbol{\varkappa}_{j'}^t}$ do not effect each other. In other words, this condition says that the impact of a turbo charger aging on a particular EAS monolith aging is neglected for instance. The input data $\boldsymbol{y}_{t,j'}^0$ of aging model $A_{j'}$ is

obtained by replicating MoBEO simulation data, gathered with a new engine environment. In particular, the simulation of "experiment n" yields, beside the observations of the considered response variables $Y_1, \ldots, Y_r$, recordings of damage causing factors for each aging model. The inputs are given as

$$
\mathcal{Y}^0_{t_n,j'} = \left(\mathbf{y}^0_{n,1_{j'}}, \ldots, \mathbf{y}^0_{n,\phi_{j'}} \right) = \begin{pmatrix} y^0_{n,1_{j'},1} & y^0_{n,2_{j'},1} & \cdots & y^0_{n,\phi_{j'},1} \\ y^0_{n,1_{j'},2} & y^0_{n,2_{j'},2} & \cdots & y^0_{n,\phi_{j'},2} \\ \vdots & \vdots & \ddots & \vdots \\ y^0_{n,1_{j'},t_n} & y^0_{n,2_{j'},t_n} & \cdots & y^0_{n,\phi_{j'},t_n} \end{pmatrix} \in \mathbb{R}^{t_n \times \phi_{j'}} \text{ for } j' = 1, \ldots, k'.
$$

$$(5.7)$$

As highlighted by figure 5.8, every simulated experiment yields different inputs for the considered aging models. After replicating the obtained input data to a unique length t, the corresponding damage statuses can be determined by the aging models and implemented. Recordings of damage causing factors generated by the resulting aged model environment shall be denoted by matrices

$$
\mathcal{Y}^t_{t_n,j'} = \left(\mathbf{y}^t_{n,1_{j'}}, \ldots, \mathbf{y}^t_{n,\phi_{j'}} \right) \in \mathbb{R}^{t_n \times \phi_{j'}} \text{ for } j' = 1, \ldots, k'. \tag{5.8}
$$

As indicated in the previous subsection, aging might affect the considered input factors $X_1, \ldots, X_k$ of the experimental design. A particular sensor accuracy, which is specified by an experiment for instance, might not be realistic after a time span t. In this case the specified sensor accuracy is overwritten by (5.4). The simulation of an "experiment n" shall be repeated for the aged engine environment so that the aging effect in terms of $Y_1, \ldots, Y_r$ becomes assessable. Thus, each experiment of the specified experimental design has to be repeated for both levels of the factor "*time*". With regard to chapter 4, two DS-designs shall be simulated during the screening procedure, one regarding the new state and the other regarding the aged state of the engine environment. Then, "*time*" corresponds to a categorical factor, whose linear effect and interaction effect with $X_1, \ldots, X_k$ can be evaluated, just as described in subsection 4.1.2, in terms of $Y_1, \ldots, Y_r$. As a consequence, the pure aging effect and its interaction with the factors of the development and usage stage becomes assessable. This enables statements such as one usage profile is more damaging as the other for example. Aging models shall be only involved, if an aging effect on one or more of the response variables is expected. If the factor "*time*" turns out to be not significant in the screening procedure, it proposed to neglect it only under special care and upon consultation with skilled experts. Just as for the screening procedure, the obtained experimental design is simulated twice, once for the new state and once for the aged state.

So far the factor "*time*" has been considered for two levels "0" and "t". The consideration of three levels "0", "t/2", "t" triples the simulation effort, as illustrated by figure 5.9. The aged state "t/2" is obtained by replicating the simulation data of the new state to a length $t/2$ and predicting the corresponding aging statuses. The input data for the aged state "t" is in turn provided by the model at point in time $t/2$. In general the simulation effort becomes κ times larger, if the factor "*time*" is considered for $\kappa > 1$ levels.

By now, it has been assumed that damage statuses do not effect each other. It is not expected however that this simplification holds in reality. Higher temperatures downstream of the turbine, which are likely caused by turbo charger aging, might have an effect on
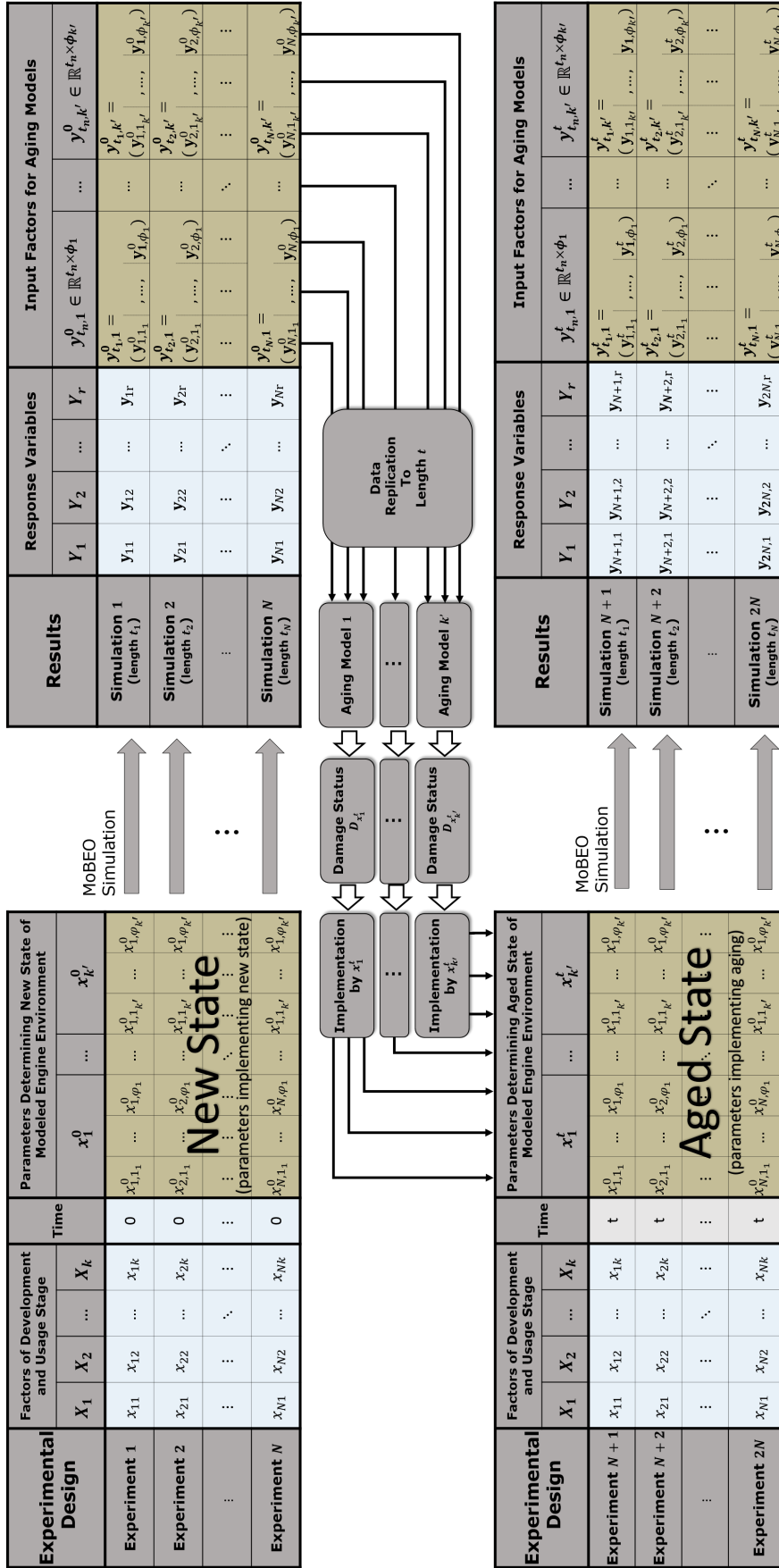
Figure 5.8: Lifetime Experimentation: Factor "*time*" at 2 Levels
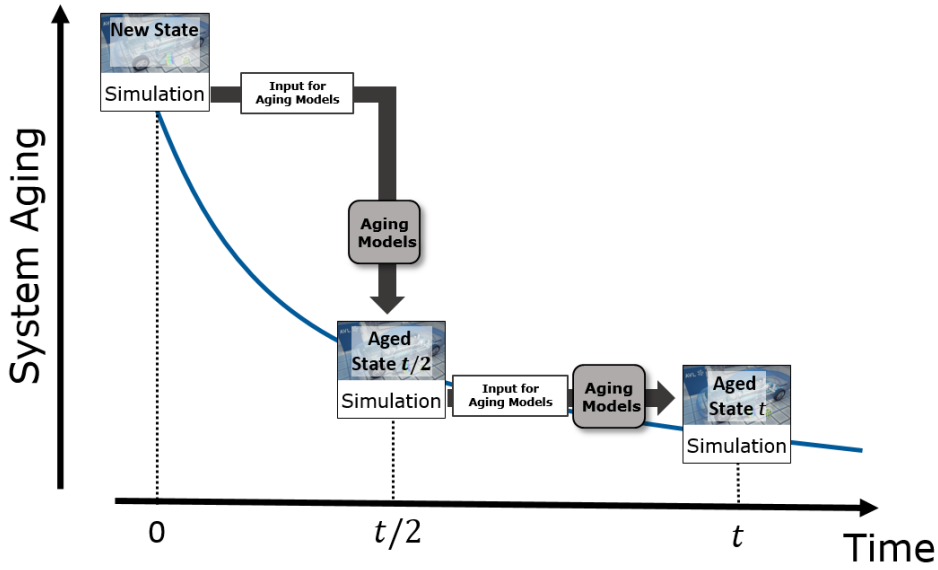
Figure 5.9: Lifetime Experimentation: Factor "*time*" at 3 Levels

the thermal aging of the DOC for instance. With respect to the air path of the engine environment, one would expect more cross influences the further back a component is located within the modeled air path. On this account, the existent simplification shall be generalized in the sense that the aging of engine components might effect the aging of EAS components. For demonstration purposes, the factor "*time*" shall be again considered for $\kappa = 2$ levels ("0" and "t"). The difference between the procedure, described in figure 5.8, and the generalized version is that the aged state "t" is considered in the first instance exclusively for engine components, as highlighted by figure 5.10. Nevertheless,
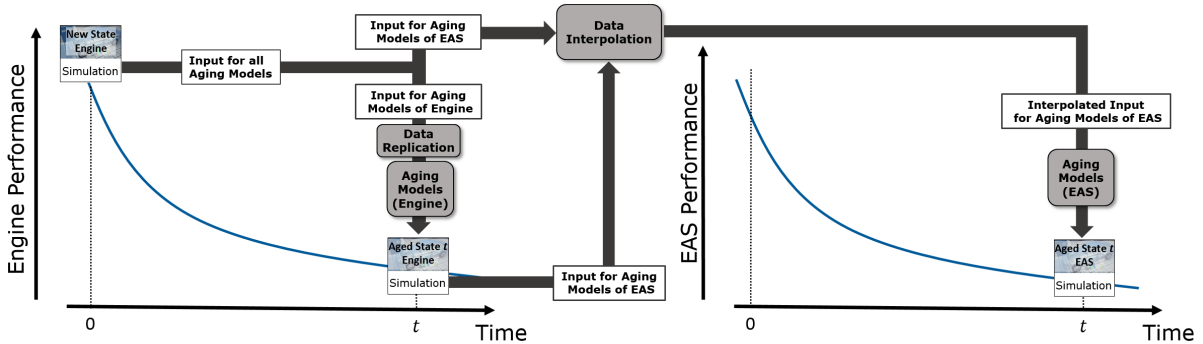
Figure 5.10: EAS Aging Depends on Engine Aging

the damage causing factors of the existing EAS aging models are recorded for both stages. Hence, $\boldsymbol{y}_{t_n,j'}^0$ and $\boldsymbol{y}_{t_n,j'}^t$ are obtained for every experiment and every aging model $A_{j'}$ of the EAS. Let $t_n > t$, then the linear interpolation yields

$$\boldsymbol{y}_{t_n,j'}^{t' \cdot t_n} = \boldsymbol{y}_{t_n,j'}^0 + \frac{\left(\boldsymbol{y}_{t_n,j'}^t - \boldsymbol{y}_{t_n,j'}^0\right) \cdot t_n}{t} \cdot t' \text{ for } t' = 0, \ldots, \lceil t/t_n \rceil. \tag{5.9}$$

The input of an EAS model $A_{j'}$ is composed of interpolated but not replicated data. The input is specifically given by

$$
\mathcal{Y}_{t,j'} = \begin{pmatrix}
\mathbf{y}_{n,1_{j'}}^{0} & \mathbf{y}_{n,2_{j'}}^{0} & \cdots & \mathbf{y}_{n,\phi_{j'}}^{0} \\
\mathbf{y}_{n,1_{j'}}^{1 \cdot t_n} & \mathbf{y}_{n,2_{j'}}^{1 \cdot t_n} & \cdots & \mathbf{y}_{n,\phi_{j'}}^{1 \cdot t_n} \\
\mathbf{y}_{n,1_{j'}}^{2 \cdot t_n} & \mathbf{y}_{n,2_{j'}}^{2 \cdot t_n} & \cdots & \mathbf{y}_{n,\phi_{j'}}^{2 \cdot t_n} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{y}_{n,1_{j'}}^{t} & \mathbf{y}_{n,2_{j'}}^{t} & \cdots & \mathbf{y}_{n,\phi_{j'}}^{t}
\end{pmatrix} \in \mathbb{R}^{t \times \phi_{j'}}. \tag{5.10}
$$

As a consequence of the interpolated input (5.10), the EAS aging eventually depends on how the engine ages. Making EAS aging dependent on engine aging triples the simulation effort, if two levels of the factor "*time*" are considered. In just the same manner, the simulation effort is increased five-fold for three levels and respectively seven-fold for four levels.

Simulation data, gathered with MoBEO, basically consists of N experiments and the consequential results of response variables $Y_1, \ldots, Y_r$. The next step is to establish a functional relationship between input factors and response variables. Particularly, *univariate regression models* shall be build for each response variable Y_s ($s = 1, \ldots, r$). Within the spanned feature space these regression models can be seen as metamodels of MoBEO, capable of simulating several thousands of experiments in a few seconds.

Chapter 6

Metamodeling Approaches and Monte Carlo Simulations

The production and lifetime-oriented development process, which was introduced in chapter 2, plays a central role in minimizing the TCO of a vehicle. On basis of the statistical concept, the RPD, the new approach allows the instant assessment of diverse quality standards in terms of given requirements and hidden costs. This is basically achieved by observing particular response variables while assessing a representative set of factor combinations regarding the development and usage stage under multiple aging states. Although the semi-physical simulation tool MoBEO is the perfect tool to simulate these factor combinations (or experiments), the simulation time available is limited in practice. The representative investigation of different feature space distributions, as defined in chapter 3, and of the considered response variables respectively is consequently not practicable. Therefore, the simulation capacity shall be rather used to set up much faster metamodels of MoBEO. The procedure on how to efficiently generate the required simulation data was subject of chapters 4 and 5. This chapter concludes the technical feasibility of the introduced production and lifetime-oriented development process by giving rise to metamodeling approaches and their application by MC simulation procedures.

The fundamental basis of metamodeling approaches will be given in section 6.1. There, the univariate regression approach, its assumptions and requirements will be outlined. Sections 6.2 and 6.3 will introduce state-of-the-art parametric and nonparametric metamodel approaches in order to precisely overcome the numerical challenge given. Section 6.4 will discuss the most important quality criteria of a metamodel, making allowance for choosing a "best" metamodel for a particular response variable. The last section 6.5 will show how a representative coverage of the feature space is derived so that quality standards become eventually assessable according to the product's life cycle.

6.1 Univariate Regression

The regression model approach is probably one of the most powerful tools in statistics. The establishing of a functional relationship between a quantitative feature and other correlated features was firstly published by the mathematicians Adrien-Marie Legendre in Legendre [1805] and by Carl Friedrich Gauss in Gauss [1809]. Udny Yule eventually

introduced the term *regression*, which originates the Latin word *regredi* (*go back or trace back to something*), into the field of statistics in Yule [1897]. The response variables $Y_1, \ldots, Y_r$ denote the quantitative features, for which such functional relationships shall be established by univariate regression models. Here, the term *univariate* highlights that one model exclusively refers to one specific response variable, denoted by Y for this section. Let $\mathbf{X}$, as given in (5.1), consist of all features that significantly correlate with response variable Y. Then, the regression model $\mathcal{R}(\mathbf{X})$ explains the variability of Y by $\mathbf{X}$ except for an additive random error ϵ (cf. Fahrmeir et al. [2010]). This is expressed by

$$Y = \mathcal{R}(\mathbf{X}) + \epsilon, \tag{6.1}$$

where it is generally assumed that the random error $\epsilon \sim N(0, \sigma^2)$. Similarly to subsection 3.1.1, the normality assumption for ϵ is argued by the CLT, where negligible influences balance on average. The n-th experiment $\mathbf{x}_n \in \mathbb{X}^k$ of an experimental design denotes a (deterministic) realization of $\mathbf{X}$. From the invariance of the normal distribution to convolution follows

$$Y_n \stackrel{\text{ind}}{\sim} N\left(\mathbb{E}\left(Y_n \mid \mathbf{x_n}\right) = \mathcal{R}\left(\mathbf{x_n}\right), \sigma^2\right) \text{ for } \mathbf{x}_n \in \mathbb{X}^k \text{ and } n = 1, \ldots, N. \tag{6.2}$$

Figure 6.1 illustrates the regression model approach. It claims that the outcome of experiment $\mathbf{x}_n$ is a realization of a normally distributed random variable Y_n with constant variance σ^2 and expected value that equals to the regression function at the point $\mathbf{x_n}$. The central regression target is to satisfy condition (6.2) with the available simulation data, given by

$$\{(y_1, \mathbf{x}_1), \ldots, (y_N, \mathbf{x}_N)\}. \tag{6.3}$$

As illustrated in figure 6.2, the objective is to find a regression function $\mathcal{R}$ and variance σ^2 so that the normal distribution of (6.2) optimally fits $(y_n, \mathbf{x}_n)$ for $n = 1, \ldots, N$. Given that
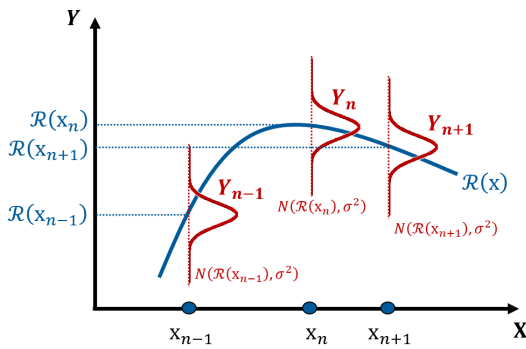
Figure 6.1: Regression Model Approach
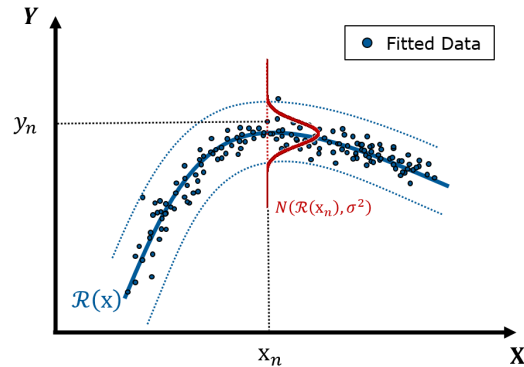
Figure 6.2: Regression Model Fit

$\mathbf{Y} = (Y_1, \ldots, Y_N)$ and $\mathbf{y} = (y_1, \ldots, y_N)$, this can be achieved by the maximum likelihood approach as follows

$$\max_{\mathcal{R}, \sigma^2} f_{(\mathbf{Y})}(\mathbf{y} \mid \mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_N), \sigma^2), \tag{6.4}$$

where

$$f_{(\mathbf{Y})}(\mathbf{y} \,|\, \mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_N), \sigma^2) = (2\pi\sigma^2)^{-(N/2)} \cdot \prod_{n=1}^{N} \exp\left(-\frac{1}{2\sigma^2} \cdot (y_n - \mathcal{R}(\mathbf{x}_n))^2\right). \qquad (6.5)$$

Maximization problem (6.4) fulfills the central regression target, as it searches for a function $\mathcal{R}$ and parameter σ^2 so that the available data optimally satisfies condition (6.2). Under the normality assumption, it can be easily shown[1] that the maximization procedure of (6.5) over $\mathcal{R}$ is independent of σ^2, and that it equals to minimizing the error function

$$\min_{\mathcal{R}} \overbrace{\frac{1}{2} \sum_{n=1}^{N} (y_n - \mathcal{R}(\mathbf{x}_n))^2}^{:=E}. \qquad (6.6)$$

Let the function $\hat{\mathcal{R}}$ denote a solution of minimization problem (6.6) that likewise satisfies

$$\hat{\mathcal{R}}(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \left(\hat{\mathcal{R}}(\mathbf{x}_1) = \hat{y}_1, \ldots, \hat{\mathcal{R}}(\mathbf{x}_N) = \hat{y}_N\right) := \hat{\mathbf{y}}. \qquad (6.7)$$

Then, maximizing (6.5) over σ^2 yields a solution, which directly depends on $\hat{\mathcal{R}}$. That solution is $\hat{\sigma}^2$, where

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} \left(y_n - \hat{\mathcal{R}}(\mathbf{x}_n)\right)^2. \qquad (6.8)$$

Under assumption (6.2), Siebertz, Van Bebber and Hochkirchen [2010] denote $\hat{\mathcal{R}}$ as a "*satisfying*" regression model, if the *residuals*

$$r_n := y_n - \hat{\mathcal{R}}(\mathbf{x}_n) = y_n - \hat{y}_n, \qquad (6.9)$$

estimating the unknown model errors $\epsilon_n \overset{\text{iid}}{\sim} N(0, \sigma^2)$, are **uncorrelated** and **zero-mean normally distributed** with **constant variance**. The procedure to check these three requirements is well known as *residual analysis*. The literature offers several procedures that can be used to check whether the residuals fulfill these requirements. Beside graphical analysis tools like autocorrelation functions, QQ-Plots or histograms, diverse statistical hypothesis tests can be applied to check the conformity of the residuals. While the *Shapiro-Wilk test*[2], the *Cramer von Mises test*[3] or the *Anderson-Darling*[4] test can be used to check whether or not the residuals origin a normal distribution, the *Breusch-Pagan test*[5], the *Goldfeld-Quandt test*[6] or the *Fligner-Killeen test*[7] might be applied to test for constant variance and the *Durbin-Watson test*[8] to detect the presence of autocorrelation. While the absence of constant variance (*heteroscedasticity*) is resolvable by the *Box-Cox*

---

[1]cf. Bishop [2007]
[2]cf. Shapiro and Wilk [1965]
[3]cf. Anderson [1962]
[4]cf. Anderson and Darling [1952]
[5]cf. Breusch and Pagan [1979]
[6]cf. Goldfeld and Quandt [1965]
[7]cf. Conover et al. [1969]
[8]cf. Durbin and Watson [1971]

transformation procedure[9], the regression model is said to be *defective*, if one or both of the other two requirements on the residuals are not fulfilled. A regression model might be defective, for instance, if a significantly correlated feature has not been taken into account by $\mathbf{X}$. For that reason, the major basis of a satisfying regression model is the involvement of expert knowledge.

Given that $\hat{\mathcal{R}}$ is a *satisfying* regression model, estimated by (6.4), the RMSE $\hat{\sigma}$ denotes the average prediction accuracy of $\hat{\mathcal{R}}$ with regard to the fitted simulation data. For all that, a small $RMSE$ is not a sufficient condition for a well predicting metamodel. As outlined in subsection 6.3.3, the $\hat{\mathcal{R}}$ can theoretically reach a complexity, where the simulation data is perfectly fitted so that

$$\hat{\mathcal{R}}(\mathbf{x}_n) = y_n \ \forall n. \tag{6.10}$$

The main purpose of a metamodel is nevertheless to predict unknown but not already simulated experiments $\mathbf{x}_* \in \mathbb{X}^k$. In fact, regression models fulfilling condition (6.10) will generally produce bad predictions for $\mathbf{x}_* \in \mathbb{X}^k$, because they reflect the random error $\epsilon \sim N(0, \sigma^2)$ but not the underlying relationship between $\mathbf{X}$ and Y. In machine learning, this phenomenon is also well known as the *over-fitting problem*. The other side of the coin is well known as the *under-fitting problem*, where the relationship between response variable and the correlated features is not totally caught. The challenge is therefore to find an acceptable trade-off between over-fitting and under-fitting, as shown in figure 6.1. Bishop [2007] proposes here to separate the available data $\{(y_1, \mathbf{x}_1), \ldots, (y_N, \mathbf{x}_N)\}$ into
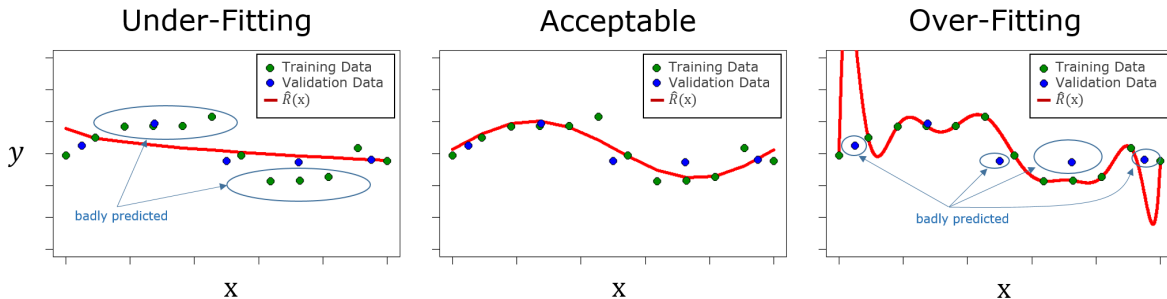
Figure 6.3: The Problem of Under-Fitting and Over-Fitting

a *training set* and into a *validation set*, whereas the validation set should exhibit a size N_{val} between $(0.01 \cdot N)$ to $(0.1 \cdot N)$. To simplicity matters, N shall denote the size of the training set for this chapter. An acceptable trade-off between over-fitting and under-fitting can be eventually achieved by observing the $RMSE$s of the training set as well as the validation set while successively increasing the *model complexity*. Ways on how to increase the model complexity depend thereby on the regression model approach chosen, and will be described in the following sections. As shown in figure 6.1, the complexity is increased as long as the $RMSE$ of the validation set decreases. At the point where the $RMSE$ of the validation set starts to ascend the procedure is stopped. It finally depends on the individual application, if the model complexity with the lowest $RMSE$

---

[9]cf. Box and Cox [1964]
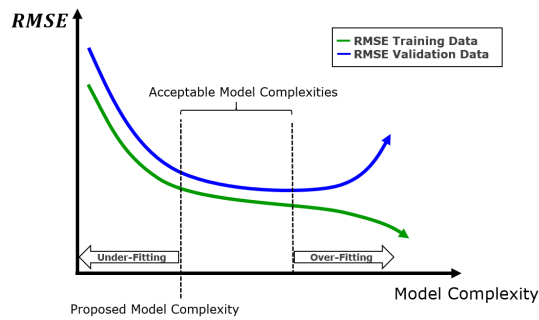
Figure 6.4: Find Optimal Metamodel Complexity by $RMSE$

regarding the validation data is chosen or not. Especially when the $RMSE$ stagnates at a certain level before starting to ascend, it is often more practical to choose the simplest possible model complexity, approximately achieving that level. Programming and using an automatized residual analysis can significantly accelerate the search for the optimal complexity.

To conclude this section, a good prediction performance regarding the validation data makes a regression model to a metamodel. Yet, it is required however to solve minimization problem (6.6). Dependent on the approach chosen for $\mathcal{R}$, the optimization procedure and controlling the model complexity must be differently overcome. Ways on how to choose $\mathcal{R}$ are generally separated into parametric and nonparametric regression approaches.

6.2 Parametric Regression Approaches

The main characteristic of parametric regression approaches is that the functional form of $\mathcal{R}$ is predefined except for variable parameter vector $\mathbf{w}$. By solving minimization problem (6.6) these are set such that the underlying data is optimally fitted. Therefore, parametric regression approaches are especially suited for metamodeling procedures, if the functional relationship between response and correlated features is assessable to a certain extend. The most famous and best researched parametric regression approach is certainly the polynomial regression (cf. Montgomery [2013] or Sachs and Hedderich [2006]).

6.2.1 Polynomial Regression

Technical experts are often capable of limiting the functional relationship to a certain polynomial order. Then, the polynomial regression approach can very well process this existing knowledge. In particular, the general functional form of the polynomial regression

model

$$\mathcal{R}(\mathbf{X}) = w_0 + \sum_{j=1}^{k} w_j \, X_j + \dots \tag{6.11}$$

$$\sum_{j=1}^{k} w_{jj} \, X_j^2 + \sum_{j<i}^{k} \sum w_{ji} \, X_j X_i + \dots \tag{6.12}$$

$$\sum_{j=1}^{k} w_{jjj} \, X_j^3 + \sum_{j<i}^{k} \sum w_{jji} \, X_j^2 X_i + \sum_{j<i}^{k} \sum w_{jii} \, X_j X_i^2 + \sum_{j<i<h}^{k} \sum\sum w_{jih} \, X_j X_i X_h + \dots, \tag{6.13}$$

comprising the variable parameter vector

$$\mathbf{w} = \big(w_1, w_2, \dots, w_k, w_{11}, w_{12} \dots, w_{kk}, w_{111}, w_{112}, \dots \big), \tag{6.14}$$

is curtailed according to the expected polynomial order. If a linear relationship is presumed for example, the employment of a first-order polynomial regression model is adequate. Thus, it is sufficient to exclusively consider term (6.11). Higher than first-order polynomial regression models are useful instead, if some amount of curvature between the response and system inputs is expected. Quadratic or cubic coherences necessitate a second-order model and a third-order model respectively. These are obtained by additionally including the terms (6.12) as well as (6.12) and (6.13) to the first-order model. It is generally dissuaded to use arbitrary large polynomial orders, because the data size N, necessary to determine $\mathbf{w}$, extensively increases with the polynomial order d of the model ("*curse of dimensionality*"). In fact, the dth-order polynomial regression model requires the estimation of

$$k_{\mathcal{R}} = 1 + \sum_{i=1}^{d} \binom{k}{i} \cdot \binom{d}{i} \tag{6.15}$$

effects. Due the general experience that most systems can be at least locally modeled by low-order polynomials, the maximum polynomial order is often restricted by (6.12) or (6.13). Once the number of effects in $\mathcal{R}$ is fixed, minimization problem (6.6) can be solved by the least-squares estimation method. Under given simulation data (6.3) and a fixed polynomial order d, the substitution of the non-linear terms by

$$x_{k+1} := x_1^2, \quad x_{2k+1} := x_1 x_2, \quad \dots, \quad x_{k_{\mathcal{R}}} := \underbrace{x_1 x_2 \dots x_d}_{d\text{-factor interaction}} \tag{6.16}$$

and

$$w_{k+1} := w_{11}, \quad w_{2k+1} := w_{12}, \quad \dots, \quad w_{k_{\mathcal{R}}} := w_{1\dots d} \tag{6.17}$$

yields a linear system of equations. This can be written down in matrix notation

$$\mathcal{R}\left(\mathbf{x}_1, \dots, \mathbf{x}_N\right) = \Phi \mathbf{w}, \tag{6.18}$$

where $\mathbf{\Phi} \in \mathbb{R}^{N \times k_{\mathcal{R}}}$ and $\mathbf{w} \in \mathbb{R}^{k_{\mathcal{R}}}$. In the regression context, the matrix

$$\mathbf{\Phi} = \begin{pmatrix} \phi(\mathbf{x}_1) \\ \phi(\mathbf{x}_2) \\ \vdots \\ \phi(\mathbf{x}_N) \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k_{\mathcal{R}}} \\ x_{21} & x_{22} & \dots & x_{2k_{\mathcal{R}}} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Nk_{\mathcal{R}}} \end{pmatrix} \in \mathbb{R}^{N \times k_{\mathcal{R}}}. \tag{6.19}$$

is also well known as the *design matrix*. If $\det\left(\mathbf{\Phi}^T\mathbf{\Phi}\right) > 0$, the least-squares estimator

$$\hat{\mathbf{w}} = \left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{y} \tag{6.20}$$

solves minimization problem (6.6) for the polynomial regression approach. If the residuals of

$$\hat{\mathcal{R}}\left(\mathbf{x}_1, \ldots, \mathbf{x}_N\right) = \mathbf{\Phi}\hat{\mathbf{w}} \tag{6.21}$$

satisfy condition (6.2), the result $y^* \in \mathbb{R}$ of an unknown experiment $\mathbf{x}_* \in \mathbb{X}^k$ can be predicted by

$$\hat{y}_* = \hat{\mathcal{R}}\left(\mathbf{x}_*\right) = \hat{\mathbf{w}}^T\phi\left(\mathbf{x}_*\right) = \sum_{i=1}^{k_{\mathcal{R}}} \hat{w}_i \cdot x_{*i}. \tag{6.22}$$

In order to prevent the final polynomial regression model from under-fitting or over-fitting, the model complexity needs to be controlled by the included effects. On this account, a two-step procedure is proposed, determining the minimal polynomial degree before the individual effects to be included. As illustrated by figure 6.5, the $RMSE$ referring to validation data is evaluated successively in the first step, starting with a first-order polynomial. In the second step, the search for the optimal model complexity is eventually refined, as shown in figure 6.6. On basis of the polynomial regression model possessing
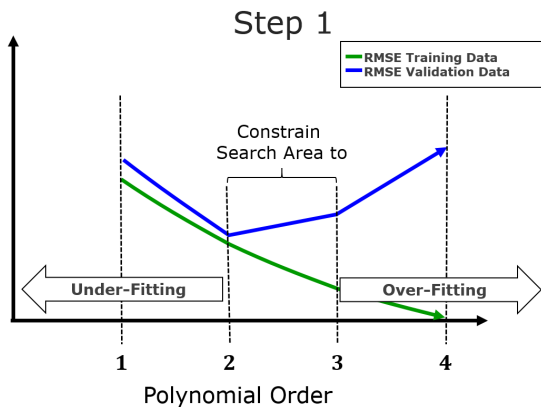
Figure 6.5: Determination of the Minimum Polynomial Order
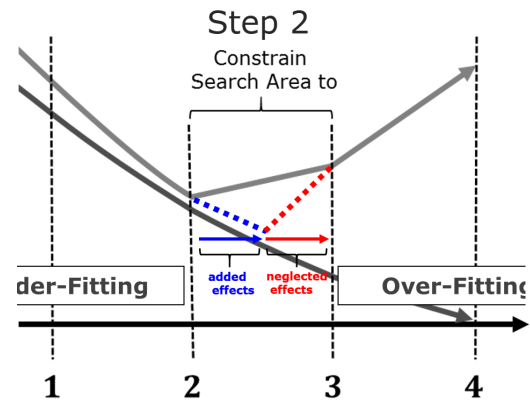
Figure 6.6: Successive Inclusion of Additional Effects

the lowest $RMSE$, this is accomplished by gradually adding all effects that belong to the next ordered model. In doing so, only those effects are effectively added that can further reduce the $RMSE$ regarding the validation data. All other possible effects are neglected. At last, the reader is reminded that every investigated regression model must pass the residual analysis before observing the $RMSE$.

6.3 Nonparametric Regression Approaches

The more complex the underlying technical system becomes, the less previous knowledge about the coherences between response and system input is usually existent, even among

experts. An optimal application of the parametric regression approach is therefore only possible to a limited extent, if at all. Nonparametric regression approaches overcome this problem, as the underlying models automatically fit the simulation data over a so called *training procedure* without providing any explicit functional form for $\mathcal{R}$.

Many nonparametric regression model approaches exist in literature. In this thesis the most notable approaches of Siebertz, Van Bebber and Hochkirchen [2010] shall be discussed, that are

- *Multivariate Adaptive Regression Splines* (MARS), cf. Friedman [1991]

- *Multilayer Perceptron* (MLP), cf. Bishop [2007]

- *Radial Basis Functions* (RBF), cf. Powell [1987]

- *Gaussian Processes* (GP), cf. Bishop [2007].

6.3.1   Multivariate Adaptive Regression Splines

Several regression spline techniques, like *B-splines*[10] or *cubic Hermite splines*[11], have been used as metamodels in industry. One of the most successful spline regression approaches has been introduced in Friedman [1991], the so called **M**ultivariate **A**daptive **R**egression **S**plines (MARS).

In practice, the curse of dimensionality makes it necessary to restrict the order of polynomial regression models. This restriction may limit, however, the global performance of the polynomial regression model, because low order polynomials are rather expected to achieve a sufficient prediction performance in local areas of the feature space (cf. Montgomery [2013]). Another limitation of the polynomial regression model is that the basis functions (6.16) are global functions in the respective features, so that the adaptation in a particular region of $\mathbb{X}^k$ affects the model performance in all other regions. Regression splines generally cope with these issues, as they split the feature space by so called *knots* into disjoint subsets that are modeled by separate but connected polynomial functions. Given a one-dimensional feature space, the spline regression approach reads as follows

$$\mathcal{R}\left(X\right) = w_0 + \sum_{i=1}^{d} w_i X^i + \sum_{m=1}^{l} w_{d+m} \left(X - K_m\right)_{+}^{d}, \tag{6.23}$$

where $\{K_1, \ldots, K_l\}$ denote the knots that split the feature space into $l+1$ subsets. As subsequently shown, (6.23) yields separate but connected polynomial functions. Expanding the regression spline approach (6.23) to k dimensions requires standardizing the underlying feature space $\mathbb{X}^k$ to $[-1, +1]^k$. Furthermore, the position of the knots is frequently chosen identically for all dimensions so that a basis function vector

$$\mathbf{s}_j\left(X\right) = \big(s_{0j}(X_j), s_{1j}(X_j), s_{2j}(X_j), \ldots, s_{dj}(X_j), s_{(d+1)j}(X_j), \ldots, s_{(d+l)j}(X_j)\big) \tag{6.24}$$

---

[10]cf. Prautzsch et al. [2001]
[11]cf. Schoenberg [1973]

1st Order Regression Splines:
cf. figure 6.7

$X < K_1:\ \mathcal{R}(X) = w_0 + w_1 X$

$X > K_1:\ \mathcal{R}(X) = w_0 + w_1 X + w_2 \left(X - K_1\right)^{d=1}$

$\qquad = \underbrace{w_0 - K_1 w_2}_{\tilde{w}_0} + \underbrace{(w_1 + w_2)}_{\tilde{w}_1} X$
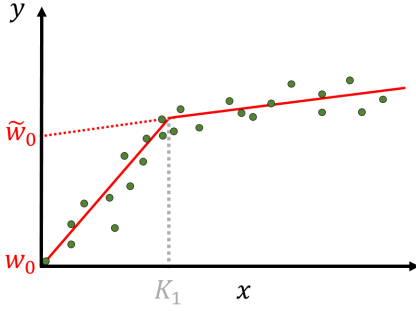
2nd Order Regression Splines:
cf. figure 6.8

$X < K_1:\ \mathcal{R}(X) = w_0 + w_1 X + w_2 X^2$

$X > K_1:\ \mathcal{R}(X) = w_0 + w_1 X + w_2 X^2 + w_3 \left(X - K_1\right)^{d=2}$

$\qquad = \underbrace{w_0 + w_3 K_1^2}_{\tilde{w}_0} + \underbrace{(w_1 - w_3 2 K_1)}_{\tilde{w}_1} X + \underbrace{(w_2 + w_3)}_{\tilde{w}_2} X^2$
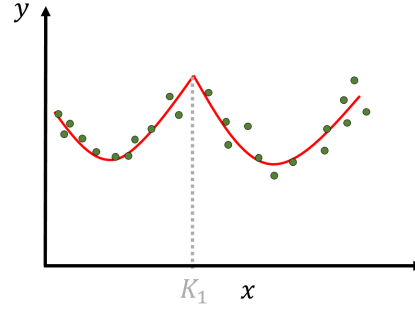
Figure 6.7: 1st Order Regression Splines

Figure 6.8: 2nd Order Regression Splines

is obtained for each considered feature X_j, $j = 1, \ldots, k$, where

$$s_{0j}(X_j) = 1, s_{1j}(X_j) = X_j, s_{2j}(X_j) = X_j^2, \ldots, s_{dj}(X_j) = X_j^d,$$
$$s_{(d+1)j}(X_j) = (X_j - K_1)_+^d, \ldots, s_{(d+l)j}(X_j) = (X_j - K_l)_+^d.$$

The multidimensional basis functions B_r result from multiplying the terms of (6.24), where each single term must correspond to a different feature j. Hence, the r-th possible combination of basis functions is given by

$$B_r\left(\mathbf{X}\right) := B_r\left(X_1, \ldots, X_k\right) = \prod_{\substack{j=1 \\ i\in\{0,\ldots,d+l\}}}^{k} s_{ij}\left(X_j\right). \tag{6.25}$$

The overall number of such basis functions B_r equals to $k_\mathcal{R} =: (d + l + 1)^k$, so that the multidimensional expansion of (6.23) can be finally written as follows.

$$\mathcal{R}(\mathbf{X}) = \sum_{r=1}^{k_\mathcal{R}} w_r B_r\left(\mathbf{X}\right). \tag{6.26}$$

For given simulation data, one may write (6.26) as

$$\mathcal{R}\left(\mathbf{x}_1, \ldots, \mathbf{x}_N\right) = \mathbf{\Phi} \mathbf{w}, \tag{6.27}$$

where

$$\mathbf{\Phi} = \begin{pmatrix} \phi\left(\mathbf{x}_1\right) \\ \phi\left(\mathbf{x}_2\right) \\ \vdots \\ \phi\left(\mathbf{x}_N\right) \end{pmatrix} = \begin{pmatrix} B_1(\mathbf{x}_1) & B_2(\mathbf{x}_1) & \ldots & B_{k_\mathcal{R}}(\mathbf{x}_1) \\ B_1(\mathbf{x}_2) & B_2(\mathbf{x}_2) & \ldots & B_{k_\mathcal{R}}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ B_1(\mathbf{x}_N) & B_2(\mathbf{x}_N) & \ldots & B_{k_\mathcal{R}}(\mathbf{x}_N) \end{pmatrix} \in \mathbb{R}^{N \times k_\mathcal{R}} \tag{6.28}$$

and $\mathbf{w} \in \mathbb{R}^{k_\mathcal{R}}$. Then, the least-squares estimator (6.20) solves minimization problem (6.6) for (6.27) and

$$\hat{\mathcal{R}}\left(\mathbf{x}_1, \ldots, \mathbf{x}_N\right) = \mathbf{\Phi}\hat{\mathbf{w}} \tag{6.29}$$

follows. Given that $\hat{\mathcal{R}}$ is a satisfying regression model with respect to the residual analysis, an unknown experiment $\mathbf{x}_* \in \mathbb{X}^k$ can be predicted by

$$\hat{y}_* = \hat{\mathcal{R}}\left(\mathbf{x}_*\right) = \hat{\mathbf{w}}^T \phi\left(\mathbf{x}_*\right) = \sum_{r=1}^{k_\mathcal{R}} \hat{w}_r \cdot B_r\left(\mathbf{x}_*\right). \tag{6.30}$$

Still, the analytical resolvability of the minimization problem is overshadowed by the fact that the number of possible basis functions exponentially increases in the number of involved features and considered knots. The estimation of

$$\mathbf{w} \in \mathbb{R}^{k_\mathcal{R}} \tag{6.31}$$

becomes therefore problematic, if at all possible ($N \overset{!}{>} k_\mathcal{R}$).

In Friedman [1991], a practically oriented version of regression splines, the MARS approach, is presented. In order to reduce the numerical complexity of the underlying model, it is proposed to restrict the valid basis functions in (6.24) to

$$s_{0j}(X_j) = 1,$$
$$s_{1j}(X_j) = (X_j - K_1)_+, \ldots, s_{lj}(X_j) = (X_j - K_l)_+,$$
$$s_{(l+1)j}(X_j) = (K_1 - X_j)_+, \ldots, s_{(2l)j}(X_j) = (X_j - K_l)_+.$$

Thus, the considered features are only allowed to linearly influence the response within the separated subsets of the feature space. The true challenge behind the MARS approach is to include adequate basis functions in (6.26) such that neither under-fitting nor over-fitting occurs. In the end, it is the uncertainty about the choice of the basis functions that makes MARS a nonparametric regression approach. Friedman [1991] proposes to conduct a two-step algorithm in this context. On basis of given simulation data (6.3), in the first step of this algorithm valid basis functions are successively added to the starting mean model

$$\hat{y} = \bar{y} \tag{6.32}$$

until a maximum number of basis functions is exceeded, or the $RMSE$ of the training data falls under a certain limit. While increasing the model complexity, such basis functions are prioritized that mostly reduce the $RMSE$ regarding the training set. In the second step, such basis functions are successively eliminated that can decrease the $RMSE$ of the validation set. Similar to the first step, functions with a high influence are chosen first.

6.3.2   Multilayer Perceptron

The term *neural network* originally roots in the search of McCulloch and Pitts [1943], Widrow and Hoff [1960] and Rosenblatt [1962] for mathematical representations regarding information processing in biological systems such as the human brain. The link to biology is however not so important anymore. Nowadays, neural networks are rather used instead

as robust, efficient and adaptive metamodels in statistics. The practically most proven neural network, the **MultiLayer Perceptron** (MLP), will be shortly introduced in this subsection.

A general neural network corresponds to a directed graph[12] G, whose vertices are denoted as *neurons*. The connecting edges exhibit real weights. Particularly, the edge directing from neuron j to neuron i comprises $w_{ij} \in \mathbb{R}$. Moreover, the value of a neuron j shall be denoted by

$$z_j = \begin{cases} z_j^{(0)} & \text{if } j \text{ is an input neuron} \\ h_j(a_j) & \text{else} \end{cases}, \tag{6.33}$$

where the quantity a_j is called *activation*. The nonlinear but differentiable function $h_j(\cdot)$ is correspondingly known as the *activation function* that is usually chosen to be the sigmoidal "tanh" or a linear function for regression applications. If j is not an input neuron, its activation is determined by

$$a_j = \sum_{i \in pre(j)} w_{ji} z_j \text{ where } pre(j) := \{i \,|\, \text{edge } (i,j) \in G\}. \tag{6.34}$$

A special case of a neural network is the *layered* neural network or the so called MLP, where neurons are classified into multiple layers, that are the input layer and one or more (usually two) hidden layers. Furthermore, each neuron of any preceding layer is connected to all neurons of the directly succeeding layer, as illustrated by figure 6.9. No other than
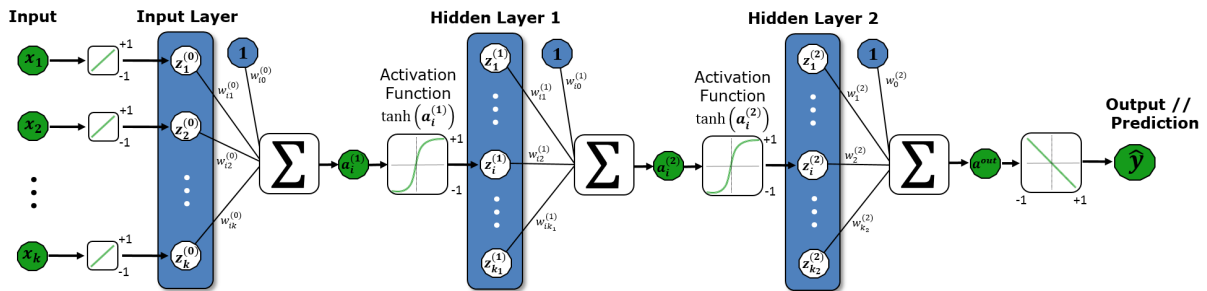
Figure 6.9: Multilayer Perceptron with 2 Hidden Layers

the connective edges just explicated are permitted for the MLP. While the values of the neurons in the input layer correspond to standardized realizations of (5.1) to $[-1, +1]$, the values of the neurons in the hidden layers result from the values of the input neurons, given weights, the chosen activation functions and (6.33). Thereby, the "tanh" as activation function covers the complete standardized range $[-1, +1]$ taken at the beginning, and the linear activation function eventually reverses the necessary standardization of the response values. The class identity is emphasized by an additional superscripted index. The hidden variables $z_0^{(0)}$, $z_0^{(1)}$ and $z_0^{(2)}$ include the so called bias parameters $w_{i0}^{(0)}$, $w_{i0}^{(1)}$ and $w_{i0}^{(2)}$.

Again, the target is to find a solution of minimization problem (6.6), given available simulation data (6.3). The minimization problem can be reformulated for the MLP as follows.

$$\min_{\mathbf{w}} E(\mathbf{w}) := \min_{\mathbf{w}} \frac{1}{2} \sum_{n=1}^{N} (y_n - \hat{y}_n(\mathbf{w}))^2. \tag{6.35}$$

---

[12]cf. Diestel [2010]

Unlike (6.20), a minimum has to be found in an iterative way, because the $\hat{y}_n(\mathbf{w})$ are generally non-linear in the network weights $\mathbf{w}$. The error function $E(\mathbf{w})$ can be approximated by the second order *Taylor polynomial* (cf. Heuser [2003]), that is

$$E(\mathbf{w} + \Delta\mathbf{w}) \approx E(\mathbf{w}) + \Delta\mathbf{w}^T \nabla_{\mathbf{w}} E(\mathbf{w}) + \frac{1}{2}\Delta\mathbf{w}^T H \Delta\mathbf{w} := \tilde{E}(\mathbf{w}) \text{ for } H := \nabla_{\mathbf{w}}^2 E(\mathbf{w}). \quad (6.36)$$

Equalizing

$$\nabla_{\Delta\mathbf{w}} \tilde{E}(\mathbf{w}) \overset{!}{=} \mathbf{0}, \quad (6.37)$$

yields the *Newton-Raphson method* with the update rule

$$\mathbf{w}(s+1) = \mathbf{w}(s) - H^{-1}\nabla_{\mathbf{w}}E(\mathbf{w}) \text{ for } s = 0, 1, 2, \ldots. \quad (6.38)$$

As long as the activation functions of the network are differentiable, the gradient $\nabla_{\mathbf{w}}E(\mathbf{w})$ can be quickly (linearly in the number of weights) determined by the *error backpropagation* procedure of Rumelhart et al. [1986]. The analytic determination of the inverse Hessian, however, might become problematic, because the computational effort necessary to determine H^{-1} cubically grows in the number of weights. In practice, the Hessian

$$H_N = \nabla_{\mathbf{w}}^2 E(\mathbf{w}) = \sum_{n=1}^{N} \nabla_{\mathbf{w}} y_n \nabla_{\mathbf{w}} y_n^T + \sum_{n=1}^{N} (y_n - \hat{y}_n(\mathbf{w})) \nabla_{\mathbf{w}} \nabla_{\mathbf{w}} y_n \quad (6.39)$$

is therefore approximated by the outer product approximation (cf. Bishop [2007])

$$H_N \approx \sum_{n=1}^{N} \nabla_{\mathbf{w}} y_n \nabla_{\mathbf{w}} y_n^T \approx \sum_{n=1}^{N} b_n b_n^T = \mathbf{J}^T \mathbf{J}, \quad (6.40)$$

where $\mathbf{J} \in \mathbb{R}^{N \times N}$ denotes the Jacobian matrix. Approximation (6.40) yields good results, if the residuals $r_n = y_n - \hat{y}_n$ are uncorrelated or $y_n \approx \hat{y}_n$. Moreover, the outer product approximation can be expanded iteratively by

$$H_{N+1} = H_N + b_{N+1}b_{N+1}^T. \quad (6.41)$$

Inserting (6.41) into the *Woodbury Identity*, given by

$$\left(M + vv^T\right)^{-1} = M^{-1} - \frac{(M^{-1}v) \cdot (v^T M^{-1})}{1 + v^T M^{-1} v}, \quad (6.42)$$

allows determining the inverse Hessian iteratively, that is

$$H_{N+1}^{-1} = H_N^{-1} - \frac{H_N^{-1} b_{N+1} b_{N+1}^T H_N^{-1}}{1 + b_{N+1}^T H_N^{-1} b_{N+1}}, \quad (6.43)$$

where $H_0 = \alpha\mathbb{I}_N$ for some α and the N-dimensional identity matrix $\mathbb{I}_N$. The approaches explained above are eventually bundled by the *Levenberg-Marquardt-backpropagation algorithm* (cf. Hagan and Menhaj [1999]). The associated update rule

$$\mathbf{w}(s+1) = \mathbf{w}(s) - \left(\mathbf{J}^T \mathbf{J} + \alpha\mathbb{I}_N\right)^{-1} \nabla_{\mathbf{w}}E(\mathbf{w}), \quad (6.44)$$

comprising an adaptive parameter α (that is chosen smaller if the target function is close to a local minimum), achieves, at least locally, quadratic convergence speed. In practice, the initial weights of (6.44) are calculated by the *Nguyen-Widrow* initialization function, described in Nguyen and Widrow [1990]. The idea behind this initialization function is to randomly set the weights in such a way that the derivative of the activation function is rather large. Although a predetermination of $\mathcal{R}$ is not required for the MLP, its basic architecture has to be defined in order to avoid under-fitting and over-fitting. Thus, before training a MLP by (6.44), an adequate number of hidden layers and associated neurons has to be figured out. In this regard, the available simulation data is repeatedly separated into a training set and into a validation set (size $\approx N \cdot 0.1$ to $N \cdot 0.2$), and the $RMSEs$ are observed for different network architectures, trained by (6.44). As a rule of thumb, it is proposed to neither use more than two layers, nor to apply as many neurons that the number of weights exceeds N. Once an acceptable network architecture has been detected, the experimenter is advised to repeat the training for different initial weights, and select the best local minimum of (6.35), which complies with the residual analysis. For a MLP with two hidden layers, as illustrated by figure 6.9, such minimum is achieved by a parameter vector $\hat{\mathbf{w}} \in \mathbb{R}^{k_{\mathcal{R}}}$ that consists of coefficients of

$$\left\{ \hat{\mathbf{w}}^{(0)}, \hat{\mathbf{w}}_0^{(0)}, \hat{\mathbf{w}}^{(1)}, \hat{\mathbf{w}}_0^{(1)}, \hat{\mathbf{w}}^{(2)}, \hat{w}_0^{(2)} \right\}, \tag{6.45}$$

where

$$\hat{\mathbf{w}}^{(0)} = \begin{pmatrix} \hat{w}_{11}^{(0)} & \hat{w}_{12}^{(0)} & \cdots & \hat{w}_{1k_1}^{(0)} \\ \hat{w}_{21}^{(0)} & \hat{w}_{22}^{(0)} & \cdots & \hat{w}_{2k_1}^{(0)} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{w}_{k1}^{(0)} & \hat{w}_{k2}^{(0)} & \cdots & \hat{w}_{kk_1}^{(0)} \end{pmatrix} \in \mathbb{R}^{k \times k_1} \text{ and } \hat{\mathbf{w}}_0^{(0)} = \begin{pmatrix} \hat{w}_{01}^{(0)} \\ \hat{w}_{02}^{(0)} \\ \vdots \\ \hat{w}_{0k_1}^{(0)} \end{pmatrix} \in \mathbb{R}^{k_1 \times 1}, \tag{6.46}$$

$$\hat{\mathbf{w}}^{(1)} = \begin{pmatrix} \hat{w}_{11}^{(1)} & \hat{w}_{12}^{(1)} & \cdots & \hat{w}_{1k_2}^{(1)} \\ \hat{w}_{21}^{(1)} & \hat{w}_{22}^{(1)} & \cdots & \hat{w}_{2k_2}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{w}_{k_11}^{(1)} & \hat{w}_{k_12}^{(1)} & \cdots & \hat{w}_{k_1k_2}^{(1)} \end{pmatrix} \in \mathbb{R}^{k_1 \times k_2} \text{ and } \hat{\mathbf{w}}_0^{(1)} = \begin{pmatrix} \hat{w}_{01}^{(1)} \\ \hat{w}_{02}^{(1)} \\ \vdots \\ \hat{w}_{0k_2}^{(1)} \end{pmatrix} \in \mathbb{R}^{k_2 \times 1}, \tag{6.47}$$

$$\hat{\mathbf{w}}^{(2)} = \begin{pmatrix} \hat{w}_1^{(1)} \\ \hat{w}_2^{(1)} \\ \vdots \\ \hat{w}_{k_2}^{(1)} \end{pmatrix} \in \mathbb{R}^{k_2 \times 1} \text{ and } \hat{w}_0^{(2)} \in \mathbb{R}. \tag{6.48}$$

Thus, it holds that

$$k_{\mathcal{R}} = k_1 \cdot (k+1) + k_2 \cdot (k_1 + 2) + 1. \tag{6.49}$$

The final regression model is given by

$$\hat{\mathcal{R}} (\mathbf{x}_1, \dots, \mathbf{x}_N) = \alpha \left(\tanh \left(\tanh \left(\mathbf{\Phi} \hat{\mathbf{w}}^{(0)} + \hat{\mathbf{w}}_0^{(0)} \right) \hat{\mathbf{w}}^{(1)} + \hat{\mathbf{w}}_0^{(1)} \right) \hat{\mathbf{w}}^{(2)} + \hat{w}_0^{(2)} \right) + \beta, \tag{6.50}$$

where the parameters $\{1/\alpha, -\beta/\alpha\}$ transform the response values $\mathbf{y}$ to $[-1, +1]$. Furthermore, the design matrix corresponds to

$$
\mathbf{\Phi} = \begin{pmatrix} \phi(\mathbf{x}_1) \\ \phi(\mathbf{x}_2) \\ \vdots \\ \phi(\mathbf{x}_N) \end{pmatrix} = \begin{pmatrix} z_{11}^{(0)} & z_{12}^{(0)} & \cdots & z_{1k}^{(0)} \\ z_{21}^{(0)} & z_{22}^{(0)} & \cdots & z_{2k}^{(0)} \\ \vdots & \vdots & \ddots & \vdots \\ z_{N1}^{(0)} & z_{N2}^{(0)} & \cdots & z_{Nk}^{(0)} \end{pmatrix} \in \mathbb{R}^{N \times k} \tag{6.51}
$$

with $z_{nj}^{(0)} = \alpha_j x_{nj} + \beta_j \in [-1, +1]$ for $n = 1, \ldots, N$. Hence, an MLP with two hidden layers predicts an unknown experiment $\mathbf{x}_* \in \mathbb{X}^k$ by

$$
\hat{y}_* = \hat{\mathcal{R}}(\mathbf{x}_*) = \alpha \left(\tanh \left(\tanh \left(\phi(\mathbf{x}_*) \hat{\mathbf{w}}^{(0)} + \hat{\mathbf{w}}_0^{(0)} \right) \hat{\mathbf{w}}^{(1)} + \hat{\mathbf{w}}_0^{(1)} \right) \hat{\mathbf{w}}^{(2)} + \hat{w}_0^{(2)} \right) + \beta, \tag{6.52}
$$

where $\phi(\mathbf{x}_*) = (\alpha_1 x_{*1} + \beta_1, \alpha_2 x_{*2} + \beta_2, \ldots, \alpha_k x_{*k} + \beta_k)$. The treatment of MLPs with more or less than two hidden layers is straightforward.

6.3.3   Radial Basis Function Networks

Radial **B**asis **F**unctions (RBF) emerge from the idea of exact data interpolation (cf. Powell [1987]). Consequently, one seeks for a functional relationship that satisfies condition (6.10). Although the exact data interpolation is not desirable for metamodels (over-fitting), a slight modification of the original RBF approach can yield well predicting metamodels.

Similar to subsections 6.2.1 and 6.3.1, let

$$
\mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \mathbf{\Phi} \mathbf{w} \tag{6.53}
$$

be a regression model that is linear in $\mathbf{w} \in \mathbb{R}^{k_{\mathcal{R}}}$. Then, the simulation data is exactly interpolated, if

$$
\mathbf{\Phi} \mathbf{w} \stackrel{!}{=} \mathbf{y} \tag{6.54}
$$

or, equivalently, if $\mathbf{\Phi}$ is a regular matrix of size $N \times N$. The more effects of multiple features are involved in a design matrix $\mathbf{\Phi}$, the more likely becomes the inclusion of redundant effects (cf. chapter 4: "*sparsity-of-effects principle*"). As a consequence, larger design matrices are usually almost singular matrices. Micchelli [1986] showed, however, that the matrix or, equivalently, if $\mathbf{\Phi}$ is a regular matrix of size $N \times N$. The more effects of multiple features are involved in a design matrix $\mathbf{\Phi}$, the more likely becomes the inclusion of redundant effects (cf. chapter 4: "*sparsity-of-effects principle*"). As a consequence, larger design matrices are usually almost singular matrices. Micchelli [1986] showed, however, that the matrix

$$
\mathbf{\Phi} = \begin{pmatrix} \phi(\mathbf{x}_1) \\ \phi(\mathbf{x}_2) \\ \vdots \\ \phi(\mathbf{x}_N) \end{pmatrix} = \begin{pmatrix} h(\|\mathbf{x}_1 - \boldsymbol{\mu}_1\|_2) & h(\|\mathbf{x}_1 - \boldsymbol{\mu}_2\|_2) & \cdots & h(\|\mathbf{x}_1 - \boldsymbol{\mu}_{k_{\mathcal{R}}}\|_2) \\ h(\|\mathbf{x}_2 - \boldsymbol{\mu}_1\|_2) & h(\|\mathbf{x}_2 - \boldsymbol{\mu}_2\|_2) & \cdots & h(\|\mathbf{x}_2 - \boldsymbol{\mu}_{k_{\mathcal{R}}}\|_2) \\ \vdots & \vdots & \ddots & \vdots \\ h(\|\mathbf{x}_N - \boldsymbol{\mu}_1\|_2) & h(\|\mathbf{x}_N - \boldsymbol{\mu}_2\|_2) & \cdots & h(\|\mathbf{x}_N - \boldsymbol{\mu}_{k_{\mathcal{R}}}\|_2) \end{pmatrix} \in \mathbb{R}^{N \times k_{\mathcal{R}}} \tag{6.55}
$$

is regular for many real valued functions h such as the *Gaussian radial basis functions*

$$\phi\left(\mathbf{x}_n\right) = h\left(\|\mathbf{x}_n - \boldsymbol{\mu}_i\|_2\right) = z_i^{(0)} := \exp\left(-\frac{\left(\|\mathbf{x}_n - \boldsymbol{\mu}_i\|_2^2\right)}{2\sigma_i^2}\right), \tag{6.56}$$

if $\mathbf{x}_n = \boldsymbol{\mu}_n$ for $n = 1, \dots, N$. Similar to regression splines, the *Gaussian radial basis functions* also overcome the disadvantage of global basis functions, as highlighted at the beginning of subsection 6.3.1. By reducing the number of basis functions $k_{\mathcal{R}}$ in (6.55), the original approach ($k_{\mathcal{R}} = N$) becomes the RBF network model, illustrated by figure 6.10. The parameters $\boldsymbol{\mu}_i$ and σ_i^2, $i = 1, \dots, k_{\mathcal{R}}$, are generally determined by the so
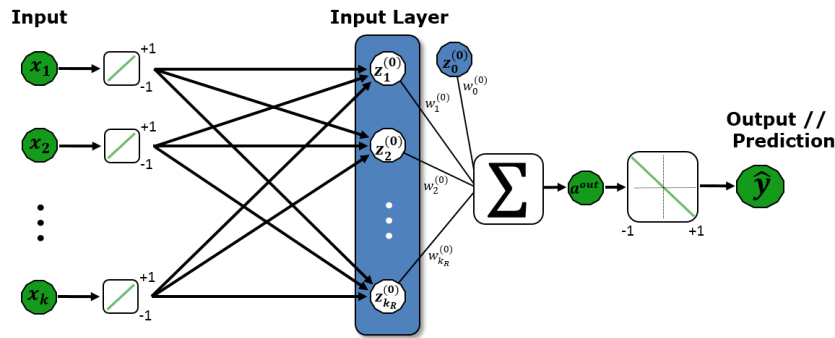
Figure 6.10: Radial Basis Function Network

called *unsupervised learning procedure* that exclusively uses input data, standardized on $[-1, +1]$. In this context, an often applied procedure is to randomly equalize the centers with standardized inputs (i.e. $\boldsymbol{\mu}_1 = \mathbf{x}_8^{\text{std}}, \boldsymbol{\mu}_2 = \mathbf{x}_N^{\text{std}}, \dots, \boldsymbol{\mu}_{k_{\mathcal{R}}} = \mathbf{x}_3^{\text{std}}$). As outlined by Bishop [2007], also other techniques on how to determine the centers $\boldsymbol{\mu}_i$ have been discussed in literature, such as *orthogonal least squares* (cf. Chen et al. [1991]), *Gaussian mixture models* (cf. Bishop [1995]) or *clustering algorithms* (cf. Moody and Darken [1989]). The treatment of the spread parameters is likely to be simplified by agreeing on $\sigma_i^2 = \sigma^2$, where the unique σ^2 must cause overlapping basis functions. A too large choice of σ^2, however, vanishes the influence of single basis functions so that the model will start to "under-fit" the relationship between $\mathbf{X}$ and Y. The spread σ^2 is typically chosen as the square of the average distance between the involved centers. Thus,

$$\sigma = \frac{1}{\binom{k_{\mathcal{R}}}{2}} \sum_{i \neq m} \|\boldsymbol{\mu}_i - \boldsymbol{\mu}_m\|_2 \cdot \delta, \tag{6.57}$$

where the adjustment factor $\delta > 1$ shall guarantee a certain amount of overlapping and that σ is smaller than the longest distance within $[-1, +1]^k$, that is

$$\sigma < \sqrt[k]{k \cdot (1 - (-1))^2}. \tag{6.58}$$

The optimal number of basis functions $k_{\mathcal{R}}$ is found by successively adding basis functions until the $RMSE$ of the training set falls beneath an error goal. Then, the number found is continuously decreased by one until the $RMSE$ of the validation set starts to increase. If the activation function is linear, the least squares estimator $\mathbf{w}^{(0)} \in \mathbb{R}^{k_{\mathcal{R}}}$, obtained by (6.20)

solves minimization problem (6.6) for the radial basis function network model. Given that the corresponding residuals comply with the residual analysis, an unknown experiment $\mathbf{x}_* \in \mathbb{X}^k$ may be predicted by

$$\hat{y}_* = \hat{\mathcal{R}}\left(\mathbf{x}_*\right) = \hat{\mathbf{w}}^T \phi\left(\mathbf{x}_*\right) = \sum_{i=1}^{k_{\mathcal{R}}} \hat{w}_i^{(0)} \cdot z_i^{(0)}, \tag{6.59}$$

where

$$z_i^{(0)} = \exp\left(-\frac{\left(\|\mathbf{x}_*^{\mathrm{std}} - \boldsymbol{\mu}_i\|_2^2\right)}{2\sigma_i^2}\right). \tag{6.60}$$

6.3.4   Kernel Methods: Gaussian Processes

All regression approaches $\mathcal{R}$ presented so far share the unique characteristic of using a vector $\mathbf{w} \in \mathbb{R}^{k_{\mathcal{R}}}$ of adaptive parameters to establish the desired functional relationship between considered features $\mathbf{X}$ and the response variable Y. In this regard, a training set of the simulation data (6.3) has been used to estimate $\mathbf{w}$ by the least-squares estimator $\hat{\mathbf{w}}$. The training set was eventually discarded, as predictions of new experiments $\mathbf{x}_* \in \mathbb{R}^k$ were exclusively based on $\hat{\mathbf{w}}$. Although these methods have the advantage that predictions are instantly generated, the determination of $\hat{\mathbf{w}}$ becomes challenging the more parameters $k_{\mathcal{R}}$ are at the bottom of $\mathbf{w}$. There exist other regression techniques, like *Gaussian processes* that overcome that disadvantage. These regression models establish an invariance to $k_{\mathcal{R}}$ by using so called *kernels*. The price to be paid, however, is the inclusion of the simulation data during the prediction phase.

The complexity of the metamodels, which have been described in the previous subsections of this chapter, is determined by the magnitude $k_{\mathcal{R}}$ of the parameter vector $\mathbf{w}$. With regard to figure 6.1, over-fitting consequently occurs, if $k_{\mathcal{R}}$ becomes too large. Bishop [2007] describes an interesting indicator for over-fitting in this context. It is stated that the more over-fitting is present, the larger become the coefficients of $\hat{\mathbf{w}}$. This principle can be adapted to the error function in (6.6) so that large vectors $\mathbf{w}$ are penalized. The modified error function is given by

$$\tilde{E}\left(\mathbf{w}\right) \coloneqq \tfrac{1}{2} \sum_{n=1}^{N} \left(y_n - \mathcal{R}\left(\mathbf{x}_n\right)\right)^2 + \tfrac{\lambda}{2}\|\mathbf{w}\|^2, \tag{6.61}$$

where the coefficient λ governs the importance of the introduced regularization. In statistics, the minimization of (6.61) is also well known as *Ridge Regression*. In matrix notation (6.61) reads as

$$\tilde{E}\left(\mathbf{w}\right) = \tfrac{1}{2}\|\boldsymbol{\Phi}\mathbf{w} - \mathbf{y}\|_2^2 + \tfrac{\lambda}{2}\|\mathbf{w}\|_2^2, \tag{6.62}$$

where $\nabla_{\mathbf{w}}\tilde{E} \overset{!}{=} \mathbf{0}$ yields

$$\hat{\mathbf{w}} = -\tfrac{1}{\lambda}\boldsymbol{\Phi}^T\left(\boldsymbol{\Phi}\mathbf{w} - \mathbf{y}\right) = \boldsymbol{\Phi}^T\hat{\mathbf{a}} \text{ with } \hat{\mathbf{a}} = -\tfrac{1}{\lambda}\left(\boldsymbol{\Phi}\mathbf{w} - \mathbf{y}\right). \tag{6.63}$$

Hence, a transition from $\hat{\mathbf{w}}$ to $\hat{\mathbf{a}}$ requires processing the input data, contained by $\boldsymbol{\Phi}$. When (6.63) is inserted into (6.62) one obtains

$$\tilde{E}\left(\mathbf{a}\right) = \tfrac{1}{2}\mathbf{a}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{a} - \mathbf{a}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{y} + \tfrac{1}{2}\mathbf{y}^T\mathbf{y} + \lambda\mathbf{a}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\mathbf{a} \tag{6.64}$$

$$= \tfrac{1}{2}\mathbf{a}^T\mathbf{K}\mathbf{K}\mathbf{a} - \mathbf{a}^T\mathbf{K}\mathbf{y} + \tfrac{1}{2}\mathbf{y}^T\mathbf{y} + \lambda\mathbf{a}^T\mathbf{K}\mathbf{a}. \tag{6.65}$$

The entries

$$K_{nm} = \phi\left(\mathbf{x}_n\right)^T \phi\left(\mathbf{x}_m\right) \coloneqq \mathcal{K}\left(\mathbf{x}_n, \mathbf{x}_m\right) \tag{6.66}$$

of the so called *Gram Matrix* $\mathbf{K} \in \mathbb{R}^{N \times N}$ are denoted as *kernels*. The least-squares estimator

$$\hat{\mathbf{a}} = \left(\mathbf{K} + \lambda \mathbb{I}_N\right)^{-1} \mathbf{y} \tag{6.67}$$

eventually results from setting $\nabla_{\mathbf{a}} E \overset{!}{=} \mathbf{0}$, and

$$\mathcal{R}\left(\mathbf{X}\right) = \hat{\mathbf{w}}^T \phi\left(\mathbf{X}\right) = \hat{\mathbf{a}}^T \boldsymbol{\Phi} \phi\left(\mathbf{X}\right) = \sum_{n=1}^{N} \hat{a}_n \cdot \mathcal{K}\left(\mathbf{X}, \mathbf{x}_n\right) \tag{6.68}$$

follows with (6.63). In machine learning, a valid kernel corresponds to a function

$$\mathcal{K} : \mathbb{X}^k \times \mathbb{X}^k \to \mathbb{R} \tag{6.69}$$

that is an inner product on $\mathbb{R}^{k_{\mathcal{R}}}$. Mercer's theorem (cf. Mercer [1909]) gives rise to the so called *kernel trick*, allowing to incorporate $k_{\mathcal{R}}$ effects in $k + 1$ arithmetic operations. The trick becomes instantly clear by the simple polynomial kernel of order $d = 2$. For two vectors $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^{k=2}$, it is given by

$$\mathcal{K}\left(\mathbf{x}, \mathbf{x}'\right) = \left(x_1 x_1' + x_2 x_2' + \beta\right)^{d=2} \tag{6.70}$$

$$= \left(x_1^2 x_1'^2 + 2x_1 x_1' x_2 x_2' + 2x_1 x_1' \beta + x_2^2 x_2'^2 + 2x_2 x_2' \beta + \beta^2\right) \tag{6.71}$$

$$= \underbrace{\left(\beta, \sqrt{2\beta}x_1, \sqrt{2\beta}x_2, \sqrt{2}x_1 x_2, x_1^2, x_2^2\right)^T}_{\phi(\mathbf{x})} \underbrace{\left(\beta, \sqrt{2\beta}x_1', \sqrt{2\beta}x_2', \sqrt{2}x_1' x_2', x_1'^2, x_2'^2\right)}_{\phi(\mathbf{x}')}.$$

$$\tag{6.72}$$

Thus, instead of computing all effects in (6.72), it suffices to simply evaluate (6.70) in order to incorporate all effects up to the quadratic order. The larger the polynomial order d is, the greater is the potential for operational savings. More generally, the *polynomial kernel* (6.70) is given for $\mathbf{x}_n, \mathbf{x}_m \in \mathbb{R}^k$ by

$$\mathcal{K}\left(\mathbf{x}_n, \mathbf{x}_m\right) = \left(\alpha \cdot \mathbf{x}_m^T \mathbf{x}_n + \beta\right)^d. \tag{6.73}$$

Karatzoglou et al. [2006] propose the application of the *ANOVA radial basis kernel*

$$\mathcal{K}\left(\mathbf{x}_n, \mathbf{x}_m\right) = \left(\sum_{j=1}^{k} \exp\left(-\sigma_{\mathcal{K}} \left(x_{nj} - x_{mj}\right)^2\right)\right)^d \tag{6.74}$$

which performs very well in multidimensional regression problems. A very well investigated regression technique in the field of kernels are Gaussian processes.

Gaussian Processes

Gaussian processes are motivated by the simple linear regression approach (6.18), which is expanded by the additional assumption that the underlying parameter vector is a normally distributed random vector $\mathbf{W}$. Consequently, one obtains

$$\mathcal{R}\left(\mathbf{X}\right) = \boldsymbol{\Phi}\mathbf{W} \sim N^N\left(\mathbf{0}, \mathbf{K}\right) \equiv \mathbb{P}\left(\mathbf{Y}_{\mathcal{R}}\right) \text{ for } \mathbf{Y}_{\mathcal{R}} \coloneqq \boldsymbol{\Phi}\mathbf{W}, \tag{6.75}$$

where $K_{nm} = \mathcal{K}(\mathbf{x}_n, \mathbf{x}_m)$. Gaussian processes become suitable for the regression approach, if the central regression assumption is adapted. Considering N experiments, assumption (6.2) can be considered under (6.75) as

$$\mathbb{P}(\mathbf{Y} \,|\, \mathbf{Y}_{\mathcal{R}} = \mathbf{\Phi}\mathbf{w}) \equiv N^N \left(\mathbf{\Phi}\mathbf{w}, \sigma^2 \mathbb{I}_N\right). \tag{6.76}$$

Then, the marginal distribution of $\mathbf{Y}$ is obtained by

$$\mathbb{P}(\mathbf{Y}) = \int \mathbb{P}(\mathbf{Y} \,|\, \mathbf{Y}_{\mathcal{R}}) \cdot \mathbb{P}(\mathbf{Y}_{\mathcal{R}}) \, d\mathbf{Y}_{\mathcal{R}} \equiv N^N (\mathbf{0}, \mathbf{C}), \tag{6.77}$$

where the matrix $\mathbf{C}$ has elements

$$\mathbf{C}_{nm} = \mathcal{K}(\mathbf{x}_n, \mathbf{x}_m) + \sigma^2 \delta_{nm}. \tag{6.78}$$

The two Gaussian sources of randomness are independent so that their covariances simply add. Furthermore, Bishop [2007] shows that

$$\mathbb{P}(\mathbf{Y}_a \,|\, \mathbf{y}_b) \equiv N^N \left(\boldsymbol{\mu}_{a|b}, \boldsymbol{\Sigma}_{a|b}\right), \tag{6.79}$$

with parameters

$$\boldsymbol{\mu}_{a|b} = \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}^{-1}(\mathbf{y}_b - \boldsymbol{\mu}_b) \tag{6.80}$$

$$\boldsymbol{\Sigma}_{a|b} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba}, \tag{6.81}$$

if

$$\mathbf{Y} = (\mathbf{Y}_a, \mathbf{Y}_b) \sim N^k (\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{6.82}$$

with parameters

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_a, \boldsymbol{\mu}_b) \ \text{and} \ \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}. \tag{6.83}$$

Result (6.79) consequently yields

$$\mathbb{P}(\mathbf{Y}_n \,|\, \mathbf{y}_{m \neq n}) \equiv N \left(\mu_{n|m}, \sigma_{n|m}^2\right), \tag{6.84}$$

where

$$\hat{\mathcal{R}}(\mathbf{x}_n) = \mu_{n|m} = \mathbf{k}_{m \neq n}^T \mathbf{C}_{N-1}^{-1} \mathbf{y}_{m \neq n} = \sum_{m \neq n}^N \hat{a}_m \cdot \mathcal{K}(\mathbf{x}_n, \mathbf{x}_m) \tag{6.85}$$

and

$$\sigma_{n|m}^2 = c - \mathbf{k}_{m \neq n}^T \mathbf{C}_{N-1}^{-1} \mathbf{k}_{m \neq n} \tag{6.86}$$

for

$$\mathbf{k}_{m \neq n} = (K_{n1}, \dots, K_{m=N,N})_{m \neq n}^T, \hat{\mathbf{a}} = \left(\mathbf{K} + \sigma^2 \mathbb{I}_N\right)^{-1} \mathbf{y} \ \text{and} \ c := K_{nn} + \sigma^2. \tag{6.87}$$

The tuning parameter σ^2 has to be optimally set in the training procedure. If the ANOVA radial basis kernel (6.74) is used, the parameter $\sigma_{\mathcal{K}}$ has to be tuned as well. In particular, Karatzoglou et al. [2006] suggest using a cross-validation procedure, where the training set is split into k_v equally sized subsets. Then, all subsets except for one are taken into

account for the generation of a *Gaussian process* so that k_v *RMSEs* in respect of the subsets left out are obtained. It is proposed to apply the parameter $\sigma_{\mathcal{K}}$, which yields the best setting in terms of the mean and standard deviation regarding *RMSE*. An unknown experiment $\mathbf{x}_* \in \mathbb{R}^k$ is finally predicted by

$$\hat{\mathbf{y}}_* = \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{y}_n = \sum_{n=1}^{N} \hat{a}_n \cdot \mathcal{K}\left(\mathbf{x}_*, \mathbf{x}_n\right), \tag{6.88}$$

where $\mathbf{k}$ and $\mathbf{C}_N$ are analogous to the upper findings.

6.4   Quality Criteria for Metamodels

In the last two subsections 6.2 and 6.3, parametric and nonparametric metamodel approaches have been introduced. The natural question is "which metamodel approach proves to be the best in modeling the semi-physical simulation tool MoBEO?".
The most intuitive target for a well performing metamodel is certainly accuracy. In general, it is desired that the metamodel predictions are as close as possible to the real simulation results of the underlying model. In engineering, a frequently used term is the *coefficient of determination* R^2, given by

Definition 6.1. (*Coefficient of Determination:* R^2)

$$R^2 = 1 - \frac{\sum_{n=1}^{N} \left(y_n - \hat{\mathcal{R}}\left(\mathbf{x}_n\right)\right)^2}{\sum_{n=1}^{N} \left(y_n - \bar{\mathbf{y}}\right)^2} \quad where \ 0 \le R^2 \le 1. \tag{6.89}$$

Hence, R^2 denotes the average prediction improvement when switching from a simple mean model $\bar{\mathbf{y}}$ to a regression model $\hat{\mathcal{R}}$. The exclusive use of R^2 regarding prediction accuracy is however not suggested for a major reason. That is, R^2 grows in the model complexity so that potential over-fitting might be not detected. This is because R^2 measures the adaptation of the model towards the training set, but not in terms of how well new experiments are predicted. Indeed, the more effects are included into the model, the more grows R^2 towards its upper boundary 1. Although that trend can be counteracted by the *adjusted coefficient of determination* R^2_{adj}, which is defined by

Definition 6.2. $\left(\boldsymbol{Adjusted\ Coefficient\ of\ Determination:}\ R^2_{adj}\right)$

$$R^2_{adj} = 1 - \frac{(N-1)}{N - k_{\mathcal{R}}} \cdot R^2 \quad where \ R^2_{adj} \le 1, \tag{6.90}$$

the prediction performance regarding new experiments is still not assessed. Montgomery [2013] suggests the application of the **PR**ediction **E**rror **S**um of **S**quares (PRESS) statistic to measure the prediction performance of a regression model. This statistic is an implicit part of the so called *Coefficient of Prediction* R^2_{Pred}

Definition 6.3. $\left(\textbf{\textit{Coefficient of Prediction:}}\ R^2_{\textbf{\textit{Pred}}}\right)$

$$R^2_{Pred} = 1 - \frac{PRESS}{\sum\limits_{n=1}^{N}\left(y_n - \bar{\mathbf{y}}\right)^2} = 1 - \frac{\sum\limits_{n=1}^{N}\left(y_n - \hat{\mathcal{R}}_{(n)}\left(\mathbf{x}_n\right)\right)^2}{\sum\limits_{n=1}^{N}\left(y_n - \bar{\mathbf{y}}\right)^2}\ for\ 0 \le R^2_{PRESS} \le 1. \qquad (6.91)$$

The model $\hat{\mathcal{R}}_{(n)}$ denotes a regression model that does not include the n-th observation for the model fit. In practice, the coefficient R^2_{Pred} is however not practical, as N regression models need to be fit in addition to the elaborate training procedure. Therefore, a more preferable solution is considering the coefficient of determination with regard to the validation data, which needs to be randomly selected in the forefront of the training procedure. Given that $N_{val} \le N$ denotes the size of the validation data, the coefficient of determination regarding the validation data is defined as

Definition 6.4. $\left(\textbf{\textit{Coefficient of Determination Aiming Validation Data:}}\ R^2_{\textbf{\textit{Val}}}\right)$

$$R^2_{Val} = 1 - \frac{\sum\limits_{n=1}^{N_{Val}}\left(y_n^{Val} - \hat{\mathcal{R}}\left(\mathbf{x}_n^{Val}\right)\right)^2}{\sum\limits_{n=1}^{N_{Val}}\left(y_n^{Val} - \bar{\mathbf{y}}^{Val}\right)^2}\ for\ 0 \le R^2_{Val} \le 1. \qquad (6.92)$$

Beside the need for prediction accuracy, a truthful inference on the feature space distribution F^k quickly requires predicting more than ten thousand experiments, as alluded in chapter 4. Thus, in addition to the prediction accuracy, a metamodel must be capable of predicting thousands of experiments within a few seconds.

The regression models introduced in this chapter have been investigated in terms of accuracy and prediction speed. Nevertheless, the performance of the metamodels differed according to the data sets. Hence, a clear identification of a "best regression approach" was not possible. This result coincides with the findings of Most and Will [2008], who had a similar experience after having compared different metamodel approaches. It is therefore generally recommended to compare different metamodel approaches in terms of (6.92) and prediction speed, and to eventually decide on the metamodel that provides the best prediction performance for the given data set.

6.5   Monte Carlo Simulation

A generated metamodel provides an accurate functional relationship between a response variable and correlated features within the boundaries of the feature space. Although this relationship is generally better (semi-physically) implemented in MoBEO, the capability to instantly derive predictions from experiments is an exclusive property of the empirical metamodel, which enables the representative investigation of a quality standard with regard to the product's life cycle. Yet the meaning and the accomplishment of such a representative investigation needs to be clarified. In this sense, the pseudo MC design is picked up again, and it is applied to metamodels to derive a distribution for the underlying

response variable. The distribution obtained can be used to determine failure probabilities and their root causes.

Let the random vector $\mathbf{X} \sim F^k$ consist of all features that significantly correlate with a response variable Y. Due to the RPD approach behind the considered production and lifetime-oriented development process, the response denotes a random variable, whose distribution F_Y only depends on the multidimensional distribution F^k of $\mathbf{X}$. Furthermore, let $\hat{\mathcal{R}}$ denote a metamodel in terms of Y and $\mathbf{X}$, which performs well inside the feature space $\mathbb{X}^k$ under considerations of this chapter such that

$$\hat{\mathcal{R}}\left(\mathbf{X}\right) \approx Y. \tag{6.93}$$

If the factor combinations

$$\{\mathbf{x}_1, \dots, \mathbf{x}_{N'}\} \text{ with } \mathbf{x}_{n'} \in \mathbb{X}^k \tag{6.94}$$

are representative for F^k, it follows from the RPD approach that

$$\{y_1, \dots, y_{N'}\} \text{ with } y_{n'} \in \mathbb{R} \tag{6.95}$$

are representative for F_Y. Thus, the determination of the unknown distribution F_Y primarily requires experiments (6.94), which are representative for F^k. From section 4.2 the reader already knows that $N' = 150.000$ experiments, generated with the LH-design of Stein [1987], are capable of representing high-dimensional normal distributions $F^k \equiv N^k\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$. As a measure of representativeness the $RMSE$ (4.41) was introduced in this context. The metamodel (6.93) allows, however, a method to assess representativeness in terms of F^k that is less cumbersome than the determination of that $RMSE$ via depth functions. Indeed, it suffices to generate pseudo MC samples $\mathbf{x}_{n'} \in \mathbb{X}^k$ of F^k and transform them into $y_{n'} \in \mathbb{R}$ by (6.93) as long as the empirical density function of Y significantly changes. Once, the empirical distribution of Y does not significantly change anymore, the generated pseudo MC samples are representative for F^k and the corresponding predictions are representative for F_Y. If multiple aging states are considered, each aging state has to be treated individually. Figure 6.11 highlights the approach on how to derive a representation of the density f_Y, whereby the two illustrated aging states differ for simplicity only by shift (the shape is generally also different). For $N' = 200$ experiments, the density f_Y clearly looks unstable and fragmentary. While the density of F_Y looks smoother, but not yet stable for $N' = 10.000$ pseudo MC samples, there is no significant difference observable when predicting $N' = 50.000$ or $N' = 150.000$ experiments by (6.93). Consequently, one can infer that the samples

$$\{\mathbf{x}_1, \dots, \mathbf{x}_{50.000}\} \text{ with } \mathbf{x}_{n'} \in \mathbb{X}^k \tag{6.96}$$

are representative for F^k. In conformity with subsection 3.1.3, one may also generate pseudo MC samples until the estimated distribution parameters do not significantly change anymore. Hence, if a normal distribution is indicated for F_Y, experiments are generated as long as the estimated parameters $\bar{y}$ and s_y^2 significantly change.

The multidimensional distribution F^k represents a certain quality standard with regard to the product's life cycle. The representative coverage of arbitrary distributions F^k, can be
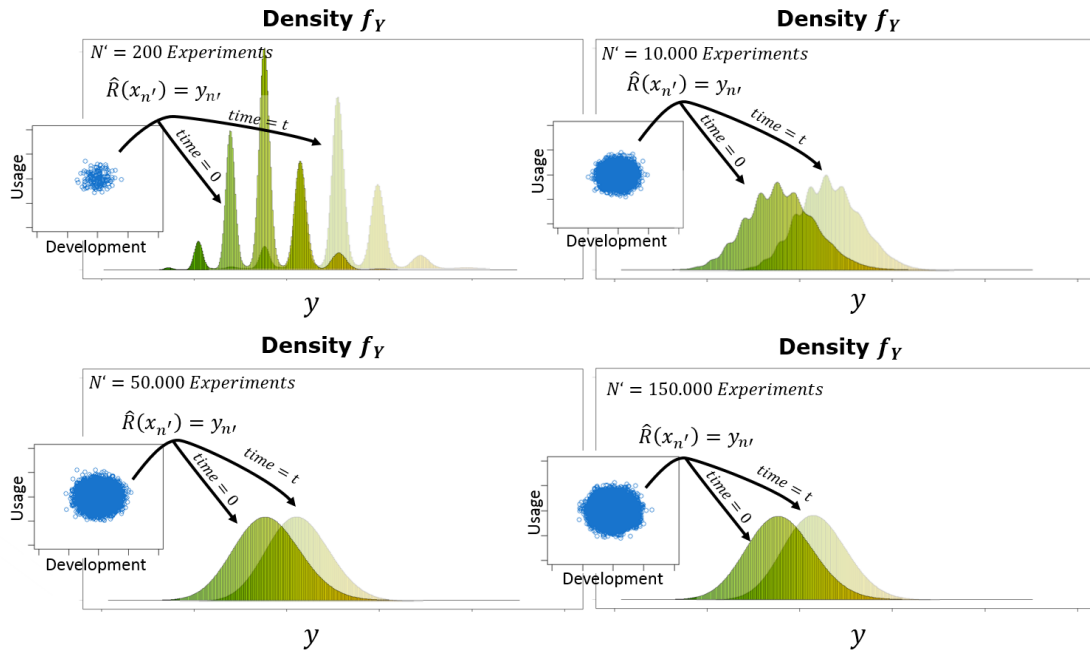
Figure 6.11: Representation of the Feature Space Distribution

achieved by the model $\hat{\mathcal{R}}\left(\mathbf{X}\right) \approx Y$, if the domains of F^k are a subset of $\mathbb{X}^k$. Consequently, the boundaries of the feature space need to be chosen always such that the investigation area is covered by $\mathbb{X}^k$. Given that two different sensor types shall be investigated for a particular application, it is always proposed to incorporate the wider specification limits in $\mathbb{X}^k$. This is because F^k is variable, while the space $\mathbb{X}^k$ is determined once at the beginning of the production and lifetime-oriented development process (cf. chapter 3). The distribution F_Y allows to instantly assess F^k in terms of how Y complies with quality requirements under the production and lifetime variability. The probability that Y exceeds a particular limit y_L for example is easily estimated by

$$\mathbb{P}\left(Y > y_L\right) := p_L \approx \frac{1}{N'} \cdot \sum_{n'=1}^{N'} \mathbf{1}_{\left\{\hat{\mathcal{R}}(\mathbf{x}_{n'}) > y_L\right\}} = \frac{1}{N'} \cdot \sum_{n'=1}^{N'} \mathbf{1}_{\left\{y'_n > y_L\right\}}. \tag{6.97}$$

As illustrated by figure 6.12, f_Y can be determined for multiple aging states in a representative manner, so that (6.97) can be evaluated respectively. Dependent on the response variable considered, this information can be processed in order to assess the compliance of particular requirements. If Y is a legally regulated emission for instance, one can directly assess the compliance with legislative requirements for the underlying feature space distribution. Another example would be that Y is a quantity, which is observed by the **On-B**oard **D**iagnostic (OBD) system. Then, (6.97) yields the probability for a light up of the corresponding MIL lamp. Furthermore, (6.97) can be used to estimate hidden costs, arising from legal penalties, warranty costs or vehicle call backs. The probability of an undesirable MIL event within the warranty period multiplied with the associated fixing costs estimates the associated warranty costs for example.

The failure probability is certainly an outstanding product of this thesis. Nevertheless, beside the evaluation of the failure probability (6.97), engineers are frequently interested
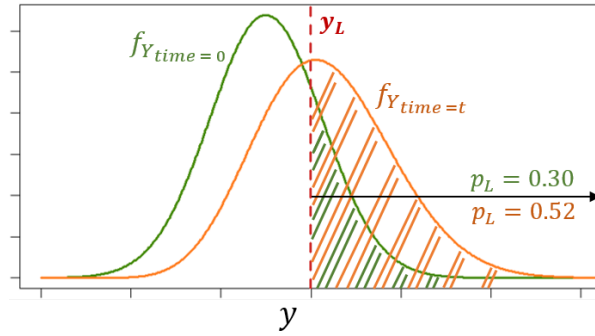
Figure 6.12: Failure Probability of a Response Variable Y

in the root causes of the failure itself. On this account, the involved factors of

$$\mathbf{X} = (X_1, \ldots, X_k) \sim F^k \tag{6.98}$$

that relate to the metamodel $\hat{\mathcal{R}}(\mathbf{X})$ shall be examined in a more detailed manner. In particular this is overcome by splitting the marginal domain Ω_j of each random variable X_j into N_j disjoint subsets or classes such that

$$\bigcup_{n_j=1}^{N_j} E_{n_j} \subseteq \Omega_j \text{ for } j = 1, \ldots, k. \tag{6.99}$$

Then, it obviously holds that

$$\mathbb{P}(Y > y_L) = \mathbb{P}\left(Y > y_L \wedge X_j \in \bigcup_{n_j=1}^{N_j} E_{n_j}\right) = \tag{6.100}$$

$$= \sum_{n_j=1}^{N_j} \mathbb{P}\left(Y > y_L \wedge X_j \in E_{n_j}\right) = \tag{6.101}$$

$$= \sum_{n_j=1}^{N_j} \mathbb{P}\left(Y > y_L \mid X_j \in E_{n_j}\right) \cdot \mathbb{P}\left(X_j \in E_{n_j}\right). \tag{6.102}$$

In this regard, the conditional probabilities

$$\mathbb{P}\left(Y > y_L \mid X_j \in E_{n_j}\right) \text{ for } n_j = 1, \ldots, N_j \tag{6.103}$$

stand in the main focus. The conditional probability (6.103) denotes the probability for a failure given that the random variable X_j realizes in a particular area of the domain Ω_j. All other involved random variables of (6.98) are thereby considered in a representative manner. The idea of this approach shall be deepened by the following example.

Example 6.1. (*Failure Probability and Root Cause Analysis*)

Let $\hat{\mathcal{R}}(\mathbf{X})$ be a metamodel in terms of a response Y, and let $\mathbf{X}$ be the random vector

$$\mathbf{X} = (X_1, X_2, X_3) \sim F^3, \tag{6.104}$$

where $F^3(\mathbf{x}) = F_{X_1, X_2}(x_1, x_2) \cdot F_{X_3}(x_3) \ \forall \mathbf{x} \in \mathbb{X}^3$. *It is assumed that the bivariate random vector*

$$(X_1, X_2) \sim N\left(\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 \\ \sigma_2\sigma_1 & \sigma_2^2 \end{pmatrix}\right) \tag{6.105}$$

incorporates the dependent factors, ambient temperature and humidity. The independent random variable

$$X_3 \sim MMULT\left(N', p_1, p_2, p_3\right) \tag{6.106}$$

involves the multinomial distributed usage behavior, considered by the usage profiles city-driving, rural-driving and highway-driving. Analyzes of the usage profiles considered resulted in the weighting parameters $p_1 = 0.60$, $p_2 = 0.25$ and $p_3 = 0.15$. The response Y denotes a counter, which causes a MIL light up when a limit of $y_L = 50$ is exceeded. In correspondence to this section, it has been found that the distribution F^3 is well represented by $N' = 50.000$ pseudo MC samples, originating F^3. Figure 6.13 illustrates the resulting density of Y, when these pseudo MC samples are applied to the metamodel given. The failure probability p_L for a MIL light up can be determined by (6.97) and equals to $p_L = 0.093$. In order to examine the root causes for running into a failure, the conditional
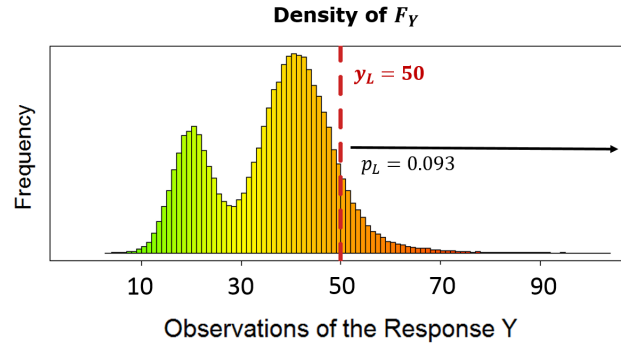
Figure 6.13: Determination of the Failure Probability

probabilities (6.103) shall be observed for X_1, X_2 and X_3. Regarding X_1, the domain Ω_1 is therefore split into N_1 disjoint classes such that

$$\overset{N_1}{\underset{n_1=1}{\bigcup}} E_{n_1} = \Omega_1. \tag{6.107}$$

Then, the conditional probability in (6.103) is approximated by

$$\mathbb{P}\left(Y > y_L \,|\, X_1 = x_{n_1} \in E_{n_1}\right) \approx \frac{1}{N'} \cdot \sum_{n'=1}^{N'} \mathbf{1}_{\{\hat{\mathcal{R}}(x_{n_1}, x_{n'2}, x_{n'3}) > y_L\}}, \tag{6.108}$$

where $x_{n'2}$ are pseudo MC samples of $F_{X_2 \,|\, X_1 = x}$ and $x_{n'3}$ of the multinomial distribution with parameters $(0.60, 0.25, 0.15)$ respectively. As illustrated in figure 6.14 for $N_1 = 100$, the conditional probability (6.108) can be illustrated in an absolute or in a relative way. The absolute representation is given by the black line, which outlines (6.108) without taking into account the underlying distribution of X_1. One can see that up to $21°C$
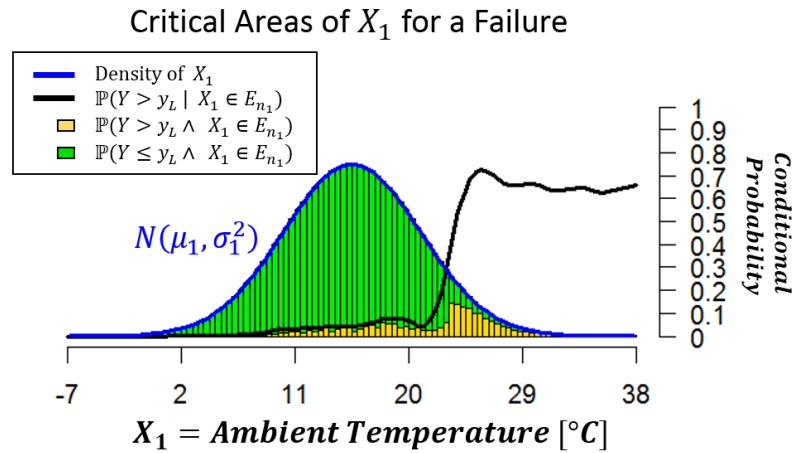
Figure 6.14: Root Cause Analysis: Ambient Temperature

the probability of Y for exceeding $y_L = 50$ is rather small, but increases rapidly above $21°C$. Thus, a MIL light up could be largely prevented, if this temperature limit is not exceeded. The key question is whether this event is probable or not. Indeed, by including the marginal distribution $N\left(\mu_1, \sigma_1^2\right)$ of X_1 into the consideration, one can see that $21°C$ is rarely exceeded. This general overview is provided by the relative consideration, which involves the different class probabilities

$$\mathbb{P}\left(X_1 \in E_{n_1}\right) = \int\limits_{E_{n_1}} \int\limits_{\Omega_2} f_{X_1,X_2}\left(x_1, x_2\right) dx_1 dx_2, \tag{6.109}$$

and so the underlying distribution of X_1. From the law of total probability, one obtains the critical and the uncritical portion of each class probability by

$$\mathbb{P}\left(X_1 \in E_{n_1}\right) = \underbrace{\mathbb{P}\left(Y > y_L \wedge X_1 \in E_{n_1}\right)}_{critical\ portion} + \underbrace{\mathbb{P}\left(Y \leq y_L \wedge X_1 \in E_{n_1}\right)}_{uncritical\ portion}. \tag{6.110}$$

In figure 6.14, the critical portions are represented by the yellow bars and the uncritical portions by the green bars. Because of (6.100) it holds that the sum of all yellow bars corresponds to the failure probability $p_L = 0.093$. The procedure on how to obtain figure 6.15 regarding X_2 is analogous to (6.108). There, it is clearly visible that ambient humidity is rather uncritical in terms of a MIL light up compared to ambient temperature. Although the absolute consideration of (6.108) outlines a large influence for low humidity values, the relative consideration indicates small probabilities for the critical subsets E_{n_2}. Moreover, the critical portions are well distributed over the considered range of humidities. Therefore, it can be concluded that the event of a MIL light up is rather independent of ambient humidity. With regard to X_3, the domain of (6.106) shall be investigated for the possible events $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$. The conditional probability in (6.103) is therefore obtained by

$$\mathbb{P}\left(Y > y_L \,|\, X_3 = x_{n_3} \in E_{n_3}\right) \approx \frac{1}{N'} \cdot \sum_{n'=1}^{N'} \mathbf{1}_{\left\{\hat{\mathcal{R}}\left(x_{n'2}, x_{n'2}, x_{n_3}\right) > y_L\right\}}. \tag{6.111}$$
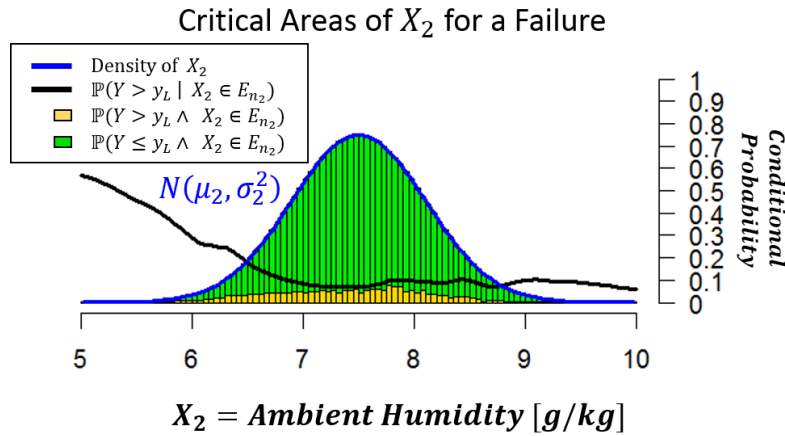
Figure 6.15: Root Cause Analysis: Ambient Humidity

The estimated conditional probabilities are illustrated by the black line in figure 6.16. While city-driving has a failure probability of approximately 34% and highway-driving of
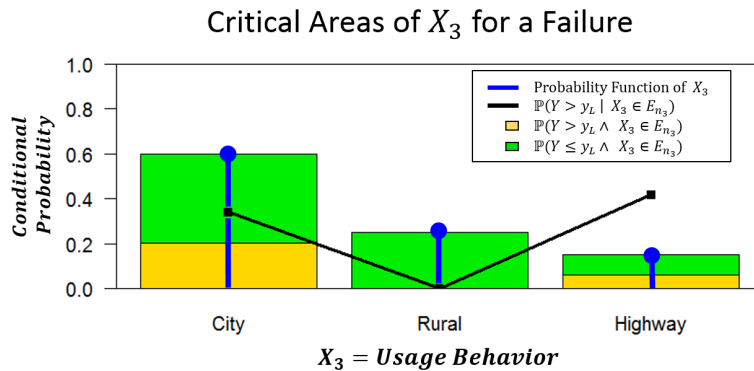
Figure 6.16: Root Cause Analysis: Usage Behavior

approximately 42%, rural driving completely prevents the observed counter from exceeding the limit $y_L = 50$. The additional observance of the class probabilities p_1, p_2 and p_3 yields critical and uncritical portions, just as described for X_1 and X_2. Overall, the absolutely more critical highway-driving turns out to be less than a half as critical as city driving. To conclude the findings, the observed counter more likely exceeds $y_L = 50$, when the vehicle is driven in cities under high temperatures. The failure probability can be finally used to determine all associated hidden costs.

Metamodeling approaches and MC simulations complete the feasibility of the introduced production and lifetime-oriented development process. The capability obtained to generate production and lifetime-representative distributions of relevant response variables raises, in fact, TCO and robustness investigations to the next level. The previous example suggested an idea of how to make use of the new process in order to quantita-

tively assess a given quality standard in terms of the usage stage. In the next chapter 7 a case study will be launched, where the new process will be applied on both stages, the development and the usage stage.

Chapter 7

Case Study - Legislative Emission Compliance

The production and lifetime-oriented development aims to achieve a solution of the central TCO minimization task, displayed in figure 2.3. Seeking for a solution of this task, the approach allows to quantitatively assess quality standards in respect of given requirements and hidden costs. It is vitally important to elaborate a process which allows to generate distributions of relevant response variables by a limited number of MoBEO experiments. After all, the above described target can be narrowed down to four work items

1. *Probabilistic Modeling & Implementation,*
2. *Design of Experiments,*
3. *Lifetime Simulations,*
4. *Metamodeling Approaches and MC Simulations.*

These topics are broadly discussed in the previous chapters, whereas this chapter presents a case study which shall illustrate how the obtained distributions can be adequately processed. Subsequently the actual implementation of the case study is revealed in detail. The target of the case study is to give the reader an insight about how an OEM could be supported with the new approach. A commonly raised issue in the automotive industry is ensuring the compliance of a vehicle or engine type with legal requirements. Among others, this concerns the assurance of an environmentally friendly system. The legal requirements apply here from the production, provided by **Conformity of Production** (CoP) requirements, to the long term use of the vehicle, set under law by **In-Service Conformity** (ISC) requirements. The reader is referred to the appendix for detailed information about CoP and ISC testing procedures. In the case study the legally regulated pollutants NO_x and NH_3 shall be checked for compliance with CoP and ISC under the integration of two different NO_x sensor types.

7.1 Test Object and Objective Target

The case study will be conducted for a heavy-duty on-road truck application with a 'technically permissible maximum laden mass' over 3.500 kg (vehicle class N_3)[1]. The vehicle is

---

[1]cf. section A.1

equipped with a **C**ompression **I**gnition (CI) engine, comprising six in line arranged cylinders that displace about 13 liters. The dual charged intake air path and the integrated common-rail fuel injection system ensure a dynamic engine response. NO_x engine out emissions are reduced by a high pressure EGR system, which recirculates a portion of the exhaust from upstream of the turbocharger to the intake manifold. A positive pressure difference between the exhaust and intake manifold as well as a better boost pressure control is enabled by two exhaust side bypasses, or so called wastegates. A sketch of the objective engine environment is given in figure 7.1. The installed EAS is supposed to
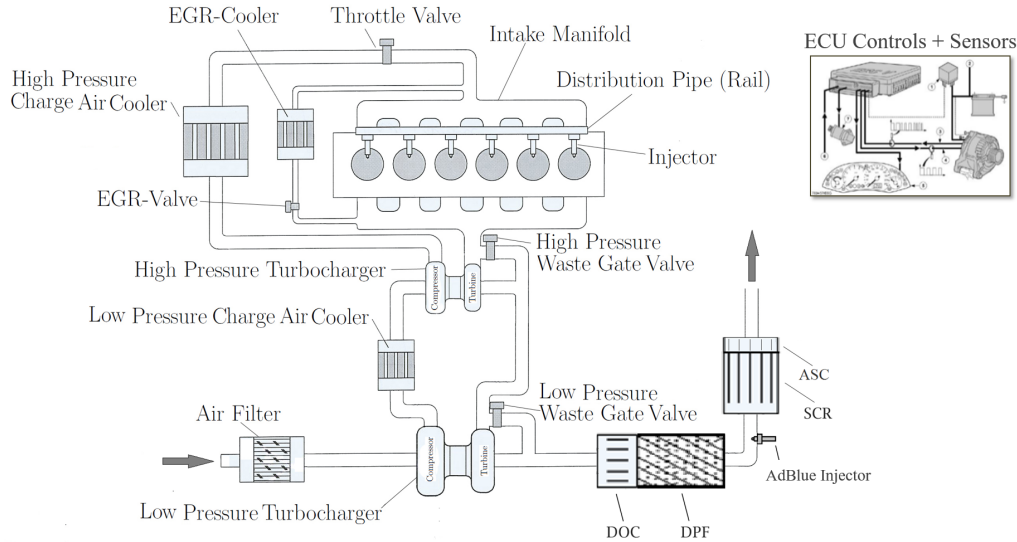
Figure 7.1: Engine Environment

fulfill the effective *EURO VI standard*, as summarized in section A.2. The EAS consists of a DOC, a DPF, a SCR and an ASC, all installed on a scale of $2.5 : 6 : 6.5 : 1$.

The objective target of this case study is the application of the production and lifetime-oriented development process in order to support the corresponding OEM in finding a satisfactory quality standard regarding the NO_x sensors, located engine out and tailpipe, to be used. On this account, two sensor types different in accuracy, produced by supplier A and B respectively, are in the main focus. The sensors are subject to the following non-linearity errors:

Supplier A: $\begin{cases} \pm 15\,ppm & NO_x < 100\,ppm \\ \pm 15\% & NO_x > 100\,ppm \end{cases}$ **Supplier B:** $\begin{cases} \pm 5\,ppm & NO_x < 100\,ppm \\ \pm 5\% & NO_x > 100\,ppm \end{cases}$

7.2   The Employed MoBEO Model

The objective engine environment is available as a highly matured MoBEO model. This means that crucial hardware components of the engine environment and their main physical characteristics were included into the model set-up. Furthermore, the model was adapted to real stationary (or *steady-state*) and dynamic (or *transient*) data, measured with the corresponding real engine environment at AVL. The widespread availability and

faster-than-real-time simulations are good reasons to decide on the MiL approach, as described in section 2.3. A realistic ECU functionality is provided by multiple closed loop controls. These concern the air system for example, acting on the amount of fresh air (controlled by the EGR valve) and the pressure inside the intake manifold (controlled by electrical waste gates). The amount of injected **Diesel Exhaust Fluid** (DEF, or commonly referred to as AdBlue$^{\text{TM}}$), added upstream SCR in order to lower the NO_x tailpipe emissions, is subject to a closed loop NH_3 storage feedback controller too. The function tolerances of the DEF injection system, that consists in most respects of the UREA dosing module, the injector nozzle, the tank and the piping, are corrected within a preconditioning run by an adaption function making use of the NO_x engine out and tailpipe sensors.

7.3   Conformity of Production

CoP has to be proofed by the manufactures after Regulation (EC) No 49 ECE. This is intended to permanently ensure the manufacturer's ability to produce a vehicle series that by and large matches the specification, performance and marking requirements as agreed with the type approval authority under the instance of manufacturing tolerances. This also concerns testing the legally regulated pollutants, NO_x and NH_3 at the exhaust tailpipe by an acceptance sampling system. Failure risks and testing effort shall be evaluated for both sensor strategies under an existent specification table for all relevant manufacturing tolerances.

The CoP testing procedure requires conducting one of the three possible acceptance sampling systems that are all described in subsection A.2.3. Whether a production series is rejected or not depends on measurement data, recorded by at least three randomly chosen unmodified engines, not longer operated than 125 hours. In the case of an on-road heavy duty application, the measurement data must origin a WHSC and WHTC testing procedure (cf. subsection A.2.1) and has to be corrected by the associated **Deterioration Factors** (DFs), as provided in subsection A.2.2. The general CoP testing procedure to be conducted is summarized by figure 7.2. CoP normally requires testing all regulated pollutants, as provided by table A.5. The testing procedure starts with three randomly chosen engines, where the legally regulated pollutants of the WHSC and WHTC are tested. The exclusive consideration of the pollutants NO_x and NH_3 is deemed sufficient for the illustration purposes of a case study. If the associated data yields neither a pass nor a rejection, additional engines (up to 32) might be tested. Assessing the failure risks and the CoP testing effort in respect of the response variables

- Y_1 ... brake specific NO_x tail pipe emissions (BSNOxTP) in g/kWh (cf. (A.1), section A.2.1),

- Y_2 ... average NH_3 ($\overline{NH_3}$) in parts per million molecules at the exhaust tailpipe,

necessitates establishing the associated distributions of Y_1 and Y_2 under the factors of the development and usage stage.
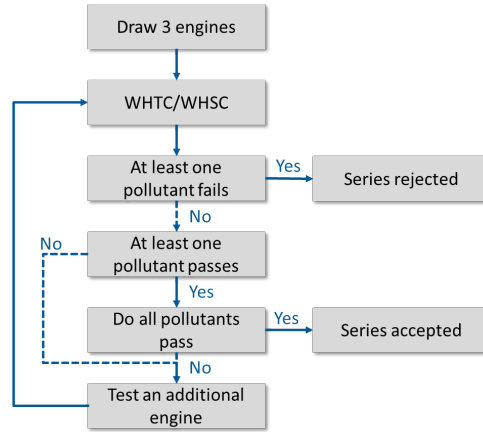
Figure 7.2: Schematic CoP Testing Procedure

In this regard, the production and lifetime-oriented development process shall be launched in order to derive these distributions. The factors of the development stage consist of one open system parameter, that is the supplier of the NO_x sensors to be used. Thus, this parameter is modeled by the discrete uniform random variable

$$U_1 \in \{\text{Supplier A}, \text{Supplier B}\}. \qquad (7.1)$$

Furthermore, a specification table of manufacturing tolerances with the most relevant parameters of MoBEO that might affect the prediction of Y_1 and Y_2 was elaborated under the considerations of section 3.1. Overall, manufacturing tolerances are considered by 45 random variables $X_1, \ldots, X_{45}$, which are listed subsequently in tables 7.2 and 7.3. Under CoP, the usage behavior is incorporated by the WHSC and WHTC, and the variable environmental conditions are deterministically set such that the normal conditions of use are fulfilled (see subsection A.2.1). Table 7.1 recapitulates the involved factors of the resulting feature space $\mathbb{X}^k$. The existent MoBEO model is used to efficiently generate
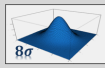
	Considered Factors		Features	Distribution Types	Impact on MoBEO
Development Stage	**Open System Parameters**		U_1	**Discrete** Uniform	**Model- or Parameter Modification**
	Manufacturing Tolerances		$X_1, \ldots, X_{45}$	**Continuous** Normal, Beta,	**Parameter Modification**
Usage Stage	**Environmental Conditions**		**25°C, 1 bar**	**Deterministic**	**Parameter Modification** (Ambient Temperature, Humidity and Pressure)
	Usage Behavior		**WHSC/WHTC**	**Deterministic**	**Usage Profile to be Simulated by Engine Speed and Torque**

Table 7.1: Case Study: CoP Factor Overview

measurement data for Y_1 and Y_2 under different factor combinations, necessary to build up most accurate metamodels. On this account, two DS-design of Jones and Nachtsheim [2011] with $N_{\text{Sc}} = 2 \cdot 45 + 1 = 91$, one for the WHSC, the other for the WHTC, are simulated.

The TSIs of Homma and Saltelli [1996], illustrated in figures 7.3 to 7.6 visualize, which factors are the most significant for Y_1 and Y_2 under WHSC and WHTC testing.
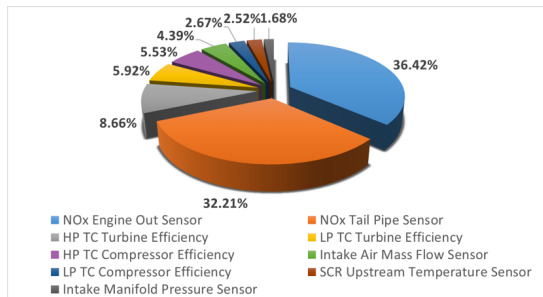
Figure 7.3: Corrected TSIs for BSNOxTP (WHSC)
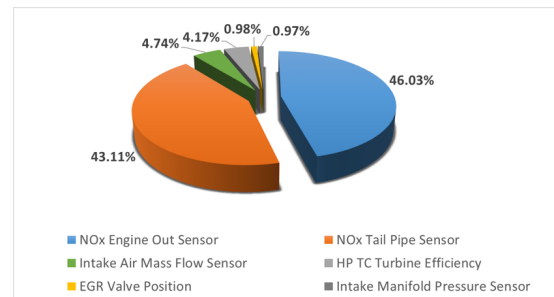
Figure 7.4: Corrected TSIs for BSNOxTP (WHTC)
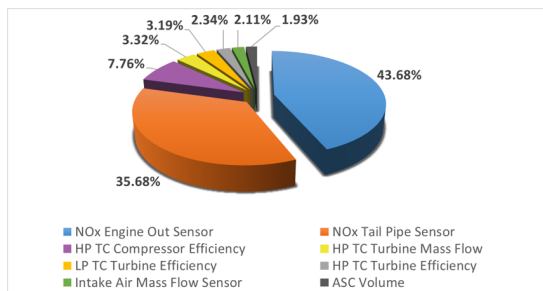
Figure 7.5: Corrected TSIs for $\overline{\text{NH}}_3$ (WHSC)
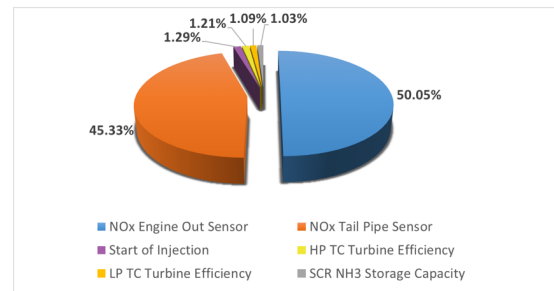
Figure 7.6: Corrected TSIs for $\overline{\text{NH}}_3$ (WHTC)

All four pie charts clearly outline that the measurement accuracy of both NO_x sensors have a significant impact on the variation of the BSNOxTP and $\overline{\text{NH}}_3$ measurement data. In fact, these results are not totally surprising, as the included adaption function regarding the DEF injection system is exclusively subject to the NO_x engine out and tailpipe sensor. Moreover, several turbo charger parameters are significant for the observed response variables. From the technical point of view this result is argued by the fact that the turbo charger has a major impact on the exhaust temperature. The exhaust temperature, in turn, significantly affects the point in time when the DEF is released upstream the SCR catalyst. The intake air mass flow sensor plays a significant role as well, because it affects the operational area of the turbo charger as a crucial part of the air mass closed loop controller.

As alluded at the beginning of subsection 4.1.4, the circumstance that Y_1 and Y_2 share almost the same significant factors brings the benefit that one LH-design can be applied for both response variables. A LH-design is generated for all significant factors of the screening procedure, and then eventually merged with a pseudo MC-design involving all other factors of tables 7.2 and 7.3. The resulting $N_{\text{Sf}} = 1500$ experiments are simulated twice, once for the WHSC and once for the WHTC. The incorporation of multiple aging

states, as discussed in chapter 5, is not scheduled for the CoP testing procedure, but implicitly considered by the legal DFs. The gathered measurement data is used to generate appropriate metamodels of the response variables for the WHSC and WHTC. In terms of chapter 6, the measurement data is randomly split into a training set and a validation set (in a ratio of 17:3) so that an optimal model complexity becomes detectable.

Actuators and Sensors

Subclass	Variation	Tolerance Interval	Random Variable
High Pressure Throttle	Effective Flap Position	$8\sigma = [-4.1\%, +4.1\%]$	$X_1 \sim F_1$ (Non-Parametric)
Common Rail System	Rail Pressure	$8\sigma = [-0.25\%, +0.25\%]$	$X_2 \sim N(\mu, \sigma^2)$
	Start of Injection	$8\sigma = [-0.51C°, +0.51C°]$	$X_3 \sim N(\mu, \sigma^2)$
EGR	Valve Position	$8\sigma = [-18.5\%, +18.5\%]$	$X_4 \sim F_4$ (Non-Parametric)
High Pressure Waste Gate	Valve Position	$8\sigma = [-30.7\%, +30.7\%]$	$X_5 \sim N(\mu, \sigma^2)$
Air Management	Intake Air Mass Flow Sensor	Gain $8\sigma = [-3\%, +3\%]$	$X_6 \sim N(\mu, \sigma^2)$
	Intake Manifold Pressure Sensor	Gain $8\sigma = [-3\%, +3\%]$	$X_7 \sim N(\mu, \sigma^2)$
Combustion Management	Pedal Position Sensor	Gain $8\sigma = [-5.0\%, +5.0\%]$	$X_8 \sim N(\mu, \sigma^2)$
	Camshaft Position Sensor	Offset $8\sigma = [-5\,rpm, +5\,rpm]$	$X_9 \sim N(\mu, \sigma^2)$
Exhaust Gas Management	Lambda Sensor	$8\sigma = [-0.01, +0.01]$ @ $\lambda = 0.8$ $8\sigma = [-0.007, +0.007]$ @ $\lambda = 1$ $8\sigma = [-0.05, +0.05]$ @ $\lambda = 1.7$	$X_{10} \sim N(\mu, \sigma^2)$
	NOx Engine Out Sensor	Supplier A Supplier B *cf. subsection 7.1*	$X_{11} \sim N(\mu, \sigma^2)$
	SCR Upstream Temperature Sensor	$8\sigma = [-3°C, +3°C]$ @ -40 to $200°C$ $8\sigma = [-1.5\%, +1.5\%]$ @ 200 to $1000°C$	$X_{12} \sim N(\mu, \sigma^2)$
	NOx Tailpipe Sensor	Supplier A Supplier B *cf. subsection 7.1*	$X_{13} \sim N(\mu, \sigma^2)$

Table 7.2: Manufacturing Tolerances: Actuators and Sensors

Engine & Exhaust Aftertreatment System

Subclass	Variation	Tolerance Interval	Random Variable
Low Pressure Turbo Charger	Compressor Efficiency	$8\sigma = [-12.6\%, +12.6\%]$	$X_{14} \sim N(\mu, \sigma^2)$
	Compressor Mass Flow	$8\sigma = [-20.0\%, +20.0\%]$	$X_{15} \sim N(\mu, \sigma^2)$
	Turbine Efficiency	$8\sigma = [-10.8\%, +10.8\%]$	$X_{16} \sim N(\mu, \sigma^2)$
	Turbine Mass Flow	$8\sigma = [-5.5\%, +5.5\%]$	$X_{17} \sim N(\mu, \sigma^2)$
Low Pressure Charge Air Cooler	Hot Effectiveness	$8\sigma = [-10\%, +10\%]$	$X_{18} \sim N(\mu, \sigma^2)$
High Pressure Turbo Charger	Compressor Efficiency	$8\sigma = [-12.6\%, +12.6\%]$	$X_{19} \sim N(\mu, \sigma^2)$
	Compressor Mass Flow	$8\sigma = [-20.0\%, +20.0\%]$	$X_{20} \sim N(\mu, \sigma^2)$
	Turbine Efficiency	$8\sigma = [-10.8\%, +10.8\%]$	$X_{21} \sim N(\mu, \sigma^2)$
	Turbine Mass Flow	$8\sigma = [-5.5\%, +5.5\%]$	$X_{22} \sim N(\mu, \sigma^2)$
High Pressure Charge Air Cooler	Hot Effectiveness	$8\sigma = [-10\%, +10\%]$	$X_{23} \sim N(\mu, \sigma^2)$
EGR Cooler	Hot Effectiveness	$8\sigma = [-1.5\%, +1.5\%]$	$X_{24} \sim N(\mu, \sigma^2)$
Diesel Oxidation Catalyst	Volume	$8\sigma = [-0.1\,l, +0.1\,l]$	$X_{25} \sim N(\mu, \sigma^2)$
	Diameter	$8\sigma = [-1.75\,mm, +1.75\,mm]$	$X_{26} \sim N(\mu, \sigma^2)$
	Intrinsic Reaction Rate	$8\sigma = [-6\%, +1\%]$	$X_{27} \sim Beta(\alpha, \beta)$
	Wall Thickness	$8\sigma = [-0.4\,mil, +0.4\,mil]$	$X_{28} \sim N(\mu, \sigma^2)$
	Cell Density	$8\sigma = [-24\,cpsi, +24\,cpsi]$	$X_{29} \sim N(\mu, \sigma^2)$
Diesel Particulate Filter	Volume	$8\sigma = [-0.2\,l, +0.2\,l]$	$X_{30} \sim N(\mu, \sigma^2)$
	Diameter	$8\sigma = [-3.5\,mm, +3.5\,mm]$	$X_{31} \sim N(\mu, \sigma^2)$
	Intrinsic Reaction Rate	$8\sigma = [-6\%, +1\%]$	$X_{32} \sim Beta(\alpha, \beta)$
	Wall Thickness	$8\sigma = [-0.4\,mil, +0.4\,mil]$	$X_{33} \sim N(\mu, \sigma^2)$
	Cell Density	$8\sigma = [-12.0\,cpsi, +12.0\,cpsi]$	$X_{34} \sim N(\mu, \sigma^2)$
Selective Catalytic Reduction	Volume	$8\sigma = [-0.2\,l, +0.2\,l]$	$X_{35} \sim N(\mu, \sigma^2)$
	Diameter	$8\sigma = [-3.5\,mm, +3.5\,mm]$	$X_{36} \sim N(\mu, \sigma^2)$
	NH3 Storage Capacity	$8\sigma = [-10\%, +10\%]$	$X_{37} \sim N(\mu, \sigma^2)$
	Wall Thickness	$8\sigma = [-0.4\,mil, 0.4\,mil]$	$X_{38} \sim N(\mu, \sigma^2)$
	Cell Density	$8\sigma = [-24\,cpsi, +24\,cpsi]$	$X_{39} \sim N(\mu, \sigma^2)$
Ammonia Slip Catalyst	Volume	$8\sigma = [-0.1\,l, +0.1\,l]$	$X_{40} \sim N(\mu, \sigma^2)$
	Diameter	$8\sigma = [-1.75\,mm, +1.75\,mm]$	$X_{41} \sim N(\mu, \sigma^2)$
	NH3 Storage Capacity	$8\sigma = [-10\%, +10\%]$	$X_{42} \sim N(\mu, \sigma^2)$
	Intrinsic Reaction Rate	$8\sigma = [-6\%, +1\%]$	$X_{43} \sim Beta(\alpha, \beta)$
	Wall Thickness	$8\sigma = [-0.4\,mil, +0.4\,mil]$	$X_{44} \sim N(\mu, \sigma^2)$
	Cell Density	$8\sigma = [-24\,cpsi, +24\,cpsi]$	$X_{45} \sim N(\mu, \sigma^2)$

Table 7.3: Manufacturing Tolerances: Engine & Exhaust Aftertreatment System

After having fit in all conscience all metamodel approaches presented in the last chapter, these shall be compared by the quality criteria, prediction accuracy and speed. Here, the prediction accuracy is captured by the coefficient of determination regarding the validation data (R^2_{Val}). Figures 7.7 to 7.10 demonstrate a clear favorite in this context, that is the MLP.
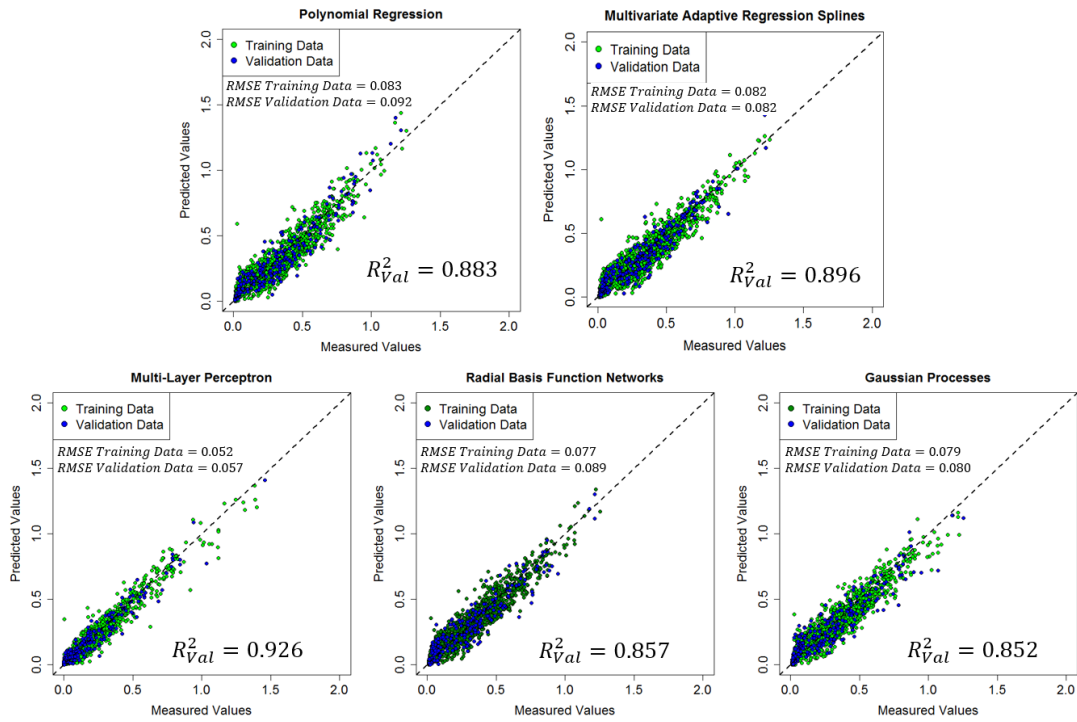
Figure 7.7: Metamodeling Approaches for BSNOxTP (WHSC), $N_{\text{Sf}} = 1.500$
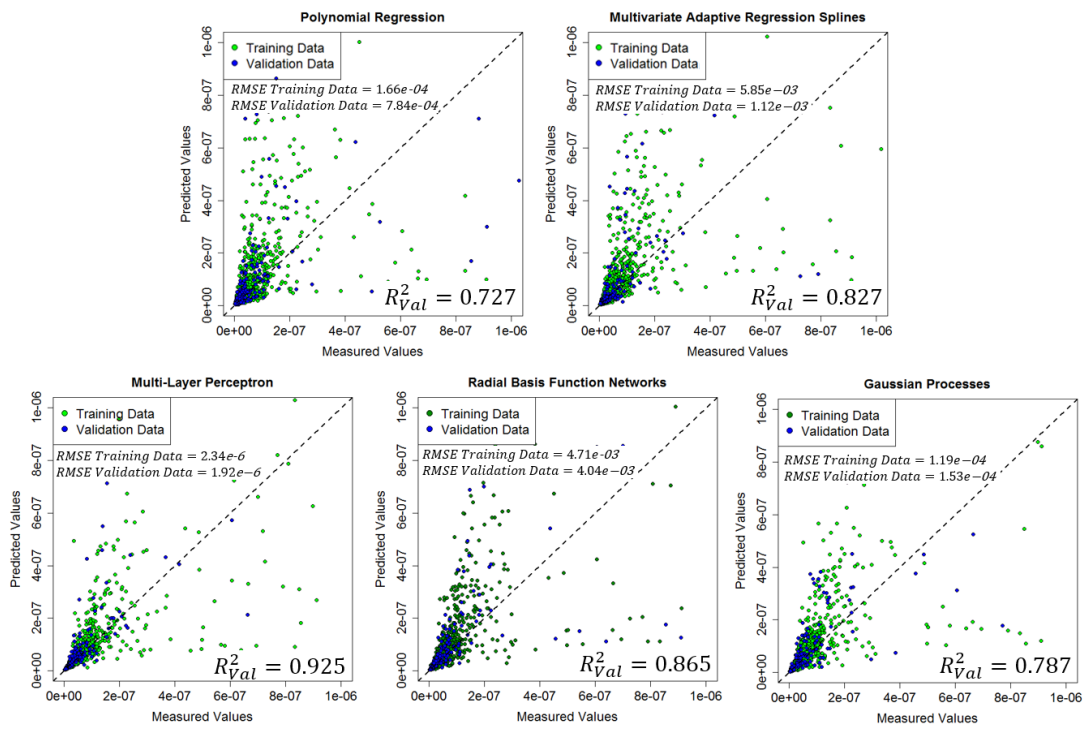
Figure 7.8: Metamodeling Approaches for $\overline{\text{NH}_3}$ (WHSC), $N_{\text{Sf}} = 1.500$
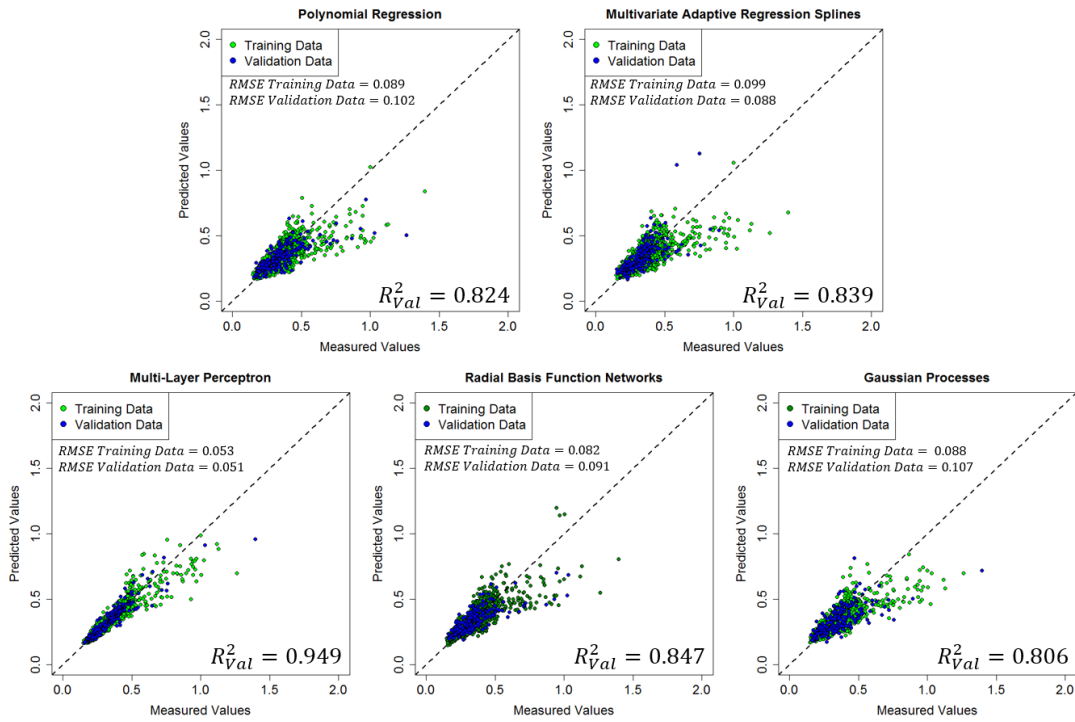
Figure 7.9: Metamodeling Approaches for BSNOxTP (WHTC), $N_{Sf} = 1.500$
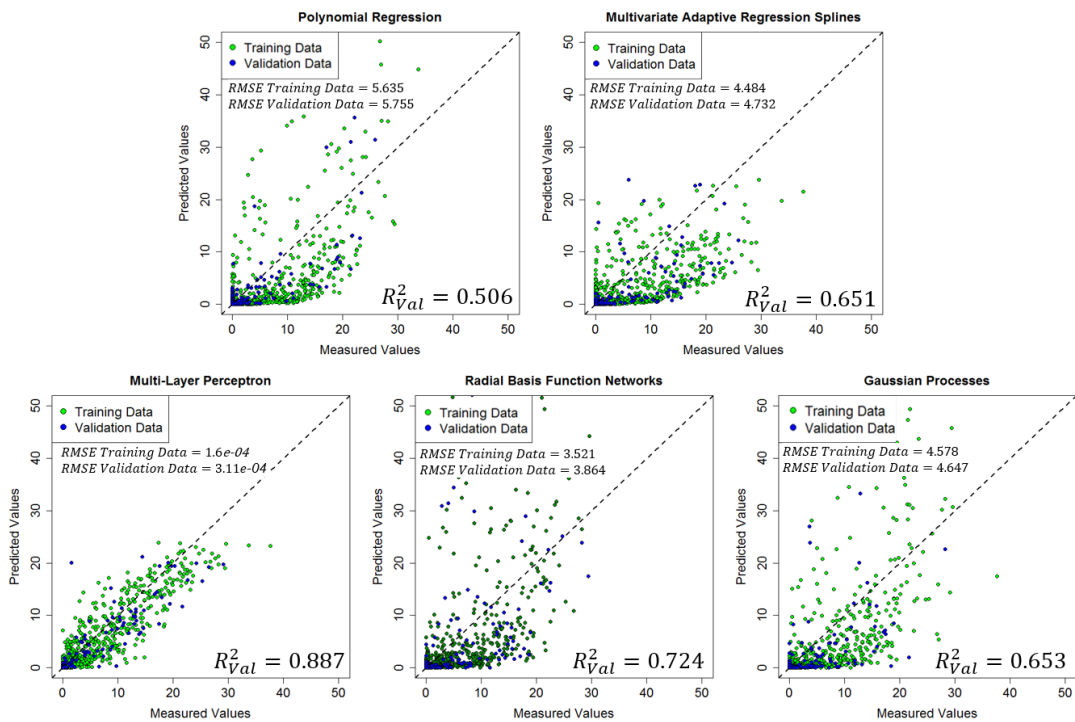
Figure 7.10: Metamodeling Approaches for $\overline{NH}_3$ (WHTC), $N_{Sf} = 1.500$

As visible in table 7.4, the MLP also scores regarding simulation speed, even if, with the exception of Gaussian Processes, the advance to its competitors is smaller than in terms of prediction accuracy. According to the results obtained, the MLP is used to

Approach	Number of Experiments	Required Time [s]
Polynomial Regression	1.00E+03	0.05
MARS	1.00E+03	0.05
MLP	1.00E+03	0.05
RBF Network	1.00E+03	0.06
Gaussian Processes	1.00E+03	24.00
Polynomial Regression	1.00E+06	2.80
MARS	1.00E+06	6.30
MLP	1.00E+06	1.50
RBF Network	1.00E+06	5.20

Table 7.4: Metamodeling Approaches: Simulation Speed

generate the distributions of BSNOxTP and $\overline{\text{NH}_3}$ for the WHSC and the WHTC under the specifications of table 7.3. In respect of section 6.5, $N' = 250.000$ pseudo MC samples prove to be enough to achieve the required representativeness regarding both NO_x sensor strategies. The corresponding distributions are illustrated by figures 7.11 and 7.12.
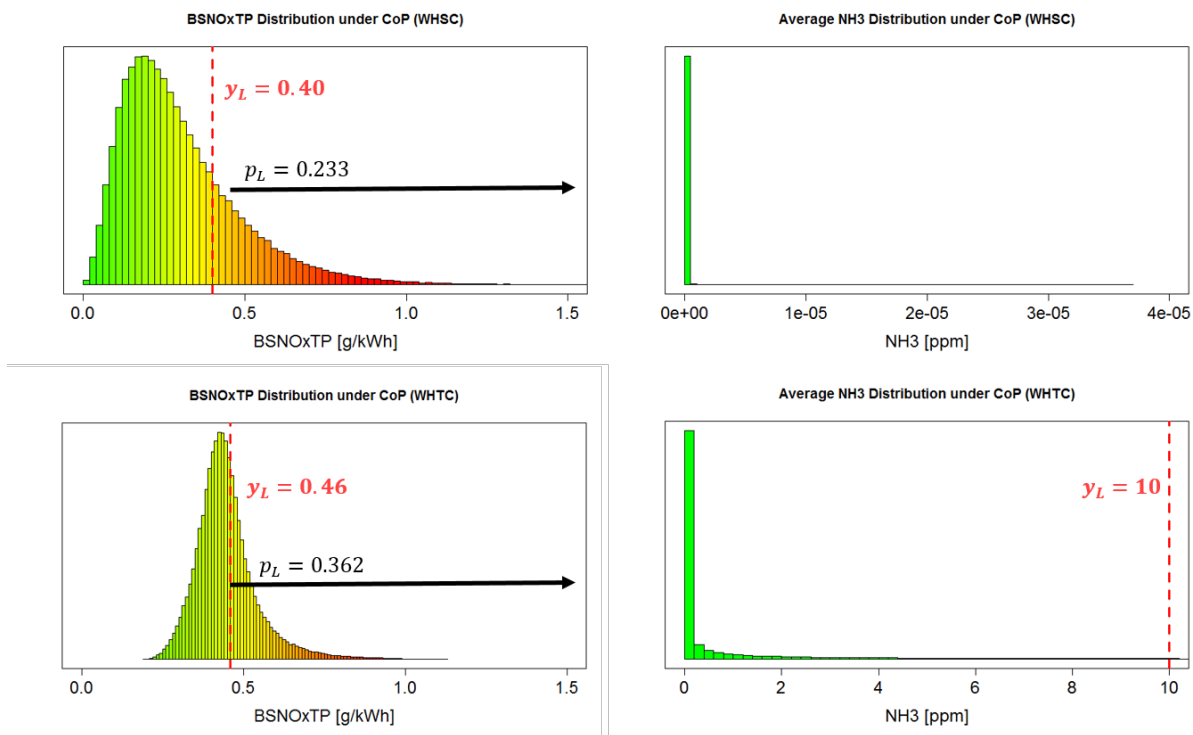
Figure 7.11: Distributions of BSNOxTP and $\overline{\text{NH}_3}$ under Supplier A
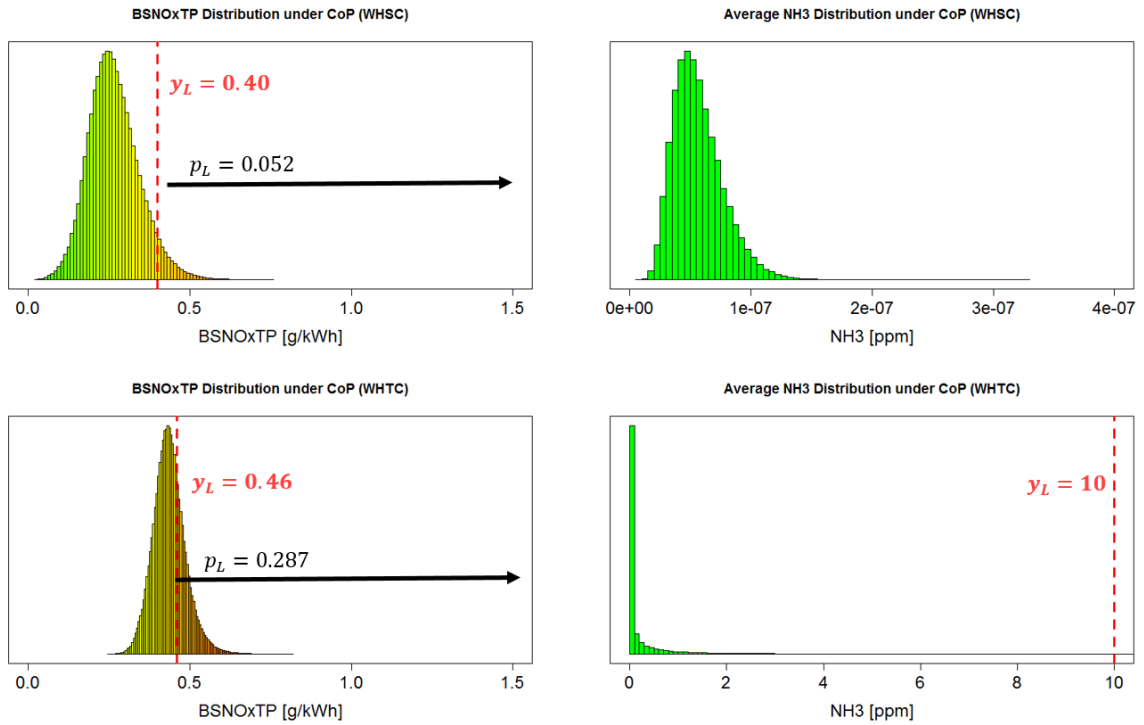
Figure 7.12: Distributions of BSNOxTP and $\overline{\text{NH}_3}$ under Supplier B

The figures shown above clearly outline that a transition from NO_x sensor supplier A to B can significantly reduce the standard deviation of all distributions. In fact, this effect is rather negligible for $\overline{\text{NH}_3}$, because the relating distributions are far away from the legislative limit of 10 ppm. Still, the transition apparently has an impact on how critical the production series presents against the legally regulated NO_x tailpipe emissions. In order to assess the associated failure risks and testing effort, the CoP testing procedure is simulated under both suppliers, whereas it is assumed that the s-method is employed. The s-method, which is described in subsection A.2.3, can be arbitrarily often repeated by taking samples of the distributions in figures 7.11 and 7.12.
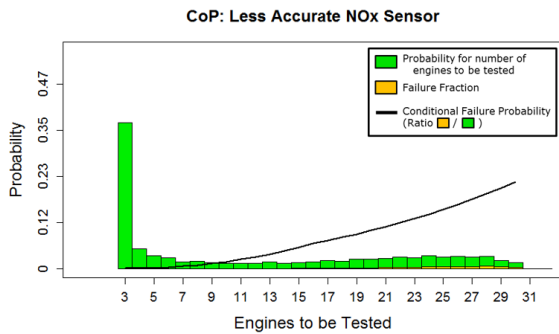
Figure 7.13: Conformity of Production: s-Method under Supplier A
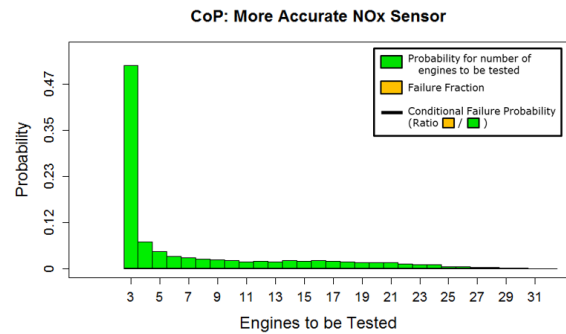
Figure 7.14: Conformity of Production: s-Method under Supplier B

The heights of the bars in figure 7.13 outline the resulting probabilities for the number of engines to be tested for the s-method when obstructing the less accurate NO_x sensors. The probability for testing three engines during the CoP test procedure is consequently approximately 36.83%, while the probability for testing four engines is already 5.07% (and so on). The color of the bars thereby emphasizes the expected outcome. The ratio of the color "yellow" within the green bars indicates the failure fraction for each number of engines to be tested. This conditional failure probability is additionally emphasized in an absolute way by the added black curve. If figure 7.13 is compared with figure 7.14, the observer easily recognizes that the probability for testing three engines is significantly higher under supplier B than under supplier A. In addition to that, while the testing procedure will be passed by all means under the more accurate sensor, this is not ensured under the less accurate sensor strategy. There it holds that the more engines are tested, the more the chances to fail CoP increase. An even better clarification of the benefit of switching to supplier B is provided by figure 7.15. There, the difference of the green bars in figure 7.14 and 7.13 are given. Thus, the agreement on the more accurate sensor type significantly
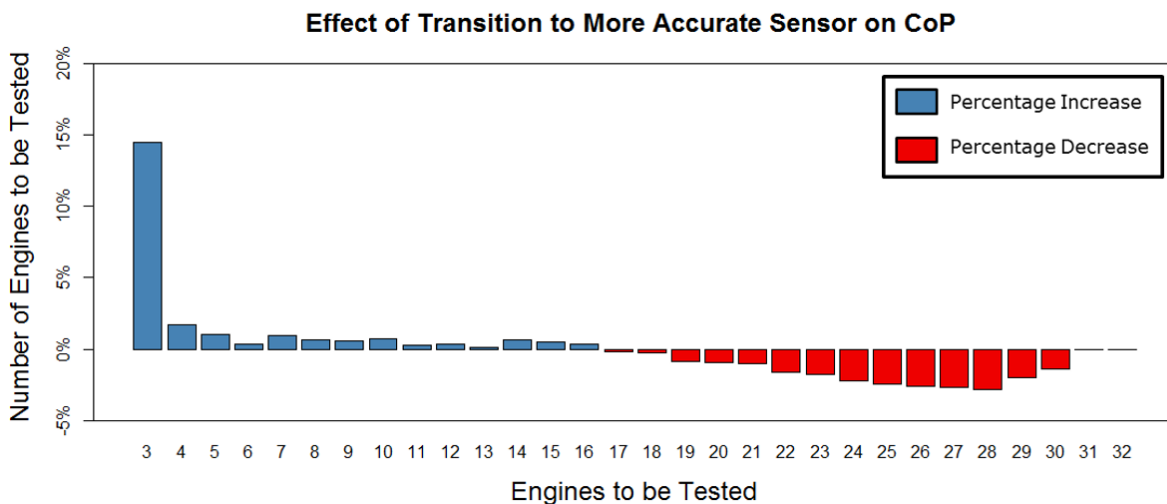
Figure 7.15: Benefit of Switching from Supplier A to Supplier B

increases the chances for testing only a few engines and simultaneously reduces the risk for testing more than 18 engines. Assuming that testing an additional engine costs 20.000$, the OEM can reduce its expected testing costs by more than 100.000$. This benefit has to be balanced by the OEM with the additional expenses when commissioning supplier B.

7.4 In-Service Conformity

Regulation (EC) No 49 ECE also includes legal standards for demonstrating (tailpipe) emission conformity of heavy-duty vehicles during their useful lives (cf. subsection A.2.2). These standards are denoted as ***In-Service Conformity*** (ISC) requirements.

The manufacturer shall consult representative test objects and record legally regulated

pollutants using **Portable Emission Measurement System** (PEMS). ISC requirements need to be shown for CO, THC, NMHC, CH_4 and NO_x. All gaseous emissions that shall be measured beside these pollutants are listed in table 1 of Regulation (EC) No 49 ECE Annex 8 Appendix 1. Representative test objects are all vehicles or engines, registered in the European Union and driven for more than 25.000 kilometers by a professional driver on real driving routs under normal load conditions. Regulation (EC) No 49 ECE Annex 8 Paragraph 4.5. defines the trip requirements as given by table 7.5. As described in

	Urban $< 50\ km/h$	Rural $\in [50,75]\ km/h$	Highway $> 75\ km/h$
M_1, N_1	45%	25%	30%
M_2, M_3	45%	25%	30%
$M_2, M_3\ (Class\ I, II, A)$	70%	30%	0%
N_2	45%	25%	30%
N_3	20%	25%	55%

Table 7.5: ISC Test Cycle Composition

subsection A.2.4, a statistical sequential sampling plan for an inspection by attributes has to be conducted in the course of the ISC testing procedure. Starting with a sample of three randomly selected test engines so called **Conformity Factors** (CFs) have to determined for all regulated pollutants listed above. The CFs are thereby determined by a method, denoted as *work based averaging window method* (cf. subsection A.2.4), which generates a series of multiple monitoring windows within the ISC test cycle. An engine is said to be *non-conforming*, if already one of the limits provided in table A.10 is exceeded by the corresponding 90% quantile of the calculated CFs. The sample is considered as

- rejected, if the number of non-conforming vehicles is larger or equal than the respective fail decision number in table A.9,

- accepted, if the number of non-conforming vehicles is smaller or equal than the respective pass decision number in table A.9.

In case neither acceptance nor rejection is achieved, an additional engine has to be consulted and evaluated correspondingly, whereas the test reaches a decision with ten engines at the latest.
Assessing ISC under table 7.3 and assumption (7.1) requires establishing the underlying distributions of the 90% quantiles of the CFs for each regulated pollutant. The introduced production and lifetime-oriented development process shall be consulted to generate such a distribution regarding the response variable

- $Y_3 \ldots 90\%$ quantile of the CFs regarding NO_x tailpipe emissions.

Again, the available MoBEO model is applied to generate the required measurement data. In contrast to the CoP testing procedure, where new engines are tested, measurement data of aged engines need to be considered instead. Unfortunately, there are no parameterized aging models available for the employed model environment. The consideration of

aging behavior by aging models, as described in chapter 5, is consequently not feasible for this case study. Nevertheless, expert knowledge of former customer projects can be used to directly implement appropriate assumptions about aging behavior. Therefore, the aging behavior of the engine is considered to simplify matters by the random variable $X_{46} \sim U(1, 1.15)$, where its realization is multiplied with the NO_x engine out emissions. A realized factor of 1.15 corresponds thereby to the legal DF of NO_x, imposed by law. Different aging modes, like thermal aging, sulfur poisoning or ash poisoning can be expected for the DOC and DPF. Aging experts believe that the catalytic activity can drop down to a residual activity of 5%. At a deeper level, a MIL light up and a consequent service of the EAS is expected. Consequently, the random variables $X_{47} \sim U(0.05, 1)$ and $X_{48} \sim U(0.05, 1)$ are introduced in order to model the aging behavior of the DOC and DPF respectively. No severe thermal aging of the SCR and ASC is expected for ISC test cycle, as the exhaust temperatures are typically below the critical limit of $500°C$. Still, the NH_3 storage capacity of the SCR and ASC may drop to approximately 50%. This issue shall be modeled by the random variables $X_{48} \sim U(0.5, 1)$ and $X_{49} \sim U(0.5, 1)$, whose realizations are multiplied with the implemented storage capacities of the SCR and ASC. Just as for the CoP testing procedure, the DS-design of Jones and Nachtsheim [2011] is applied on the resulting feature space $\mathbb{X}^k$, spanned by the factors provided by table 7.6. Thus, $N_{Sc} = 2 \cdot 49 + 1 = 99$ experiments are simulated for the screening procedure, whereas

	Considered Factors	Features	Distribution Types	Impact on MoBEO
Development Stage	**Open System Parameters**	U_1	**Discrete** Uniform	**Model- or Parameter Modification**
	Manufacturing Tolerances	$X_1, ..., X_{45}$	**Continuous** Normal, Beta,	**Parameter Modification**
Usage Stage	**Environmental Conditions**	25°C, 1 bar	**Deterministic**	**Parameter Modification** (Ambient Temperature, Humidity and Pressure)
	Usage Behavior	ISC	**Deterministic**	**Usage Profile to be Simulated by Engine Speed and Torque**
System Deterioration	**Aging Behavior** (no Aging Models available)	$X_{46}, ..., X_{49}$	**Continuous** Uniform	**DF applied on EO emissions DOC & DPF Aging SCR & ASC Storage Capacity**

Table 7.6: Case Study: ISC Factor Overview

the 90%-quantiles of the obtained CFs regarding NO_x tailpipe emissions are recorded for each experiment. The pie plot in figure 7.16 visualizes the TSIs of the most significant factors involved. Once again, the NO_x sensors turn out to be the most significant factors for the response variable considered. Beside the NO_x sensors there are several other crucial factors regarding Y_3, that are the intake air mass flow sensor, the SCR wall thickness, the high pressure turbo charger compressor efficiency, the EGR cooler efficiency and the start of injection. In contrast, the additionally considered factors $X_{46}, \ldots, X_{49}$ in order to incorporate the system aging behavior have no significant impact on Y_3. According to chapter 4, a LH-design with $N_{Sf} = 500$ is generated for the significant factors of figure 7.16 and merged with a MC-design considering all other factors of $\mathbb{X}^k$. Just as for the CoP testing procedure, the obtained data is split into a training and validation set (in
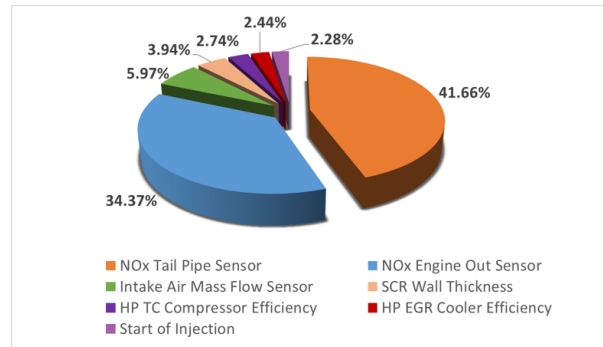
Figure 7.16: Corrected TSIs of Y_3

a ratio of 17:3), and the discussed metamodel approaches are applied in all conscience under consideration of chapter 6. The scatter plots in figure 7.17 and the respective coefficients of determination regarding the validation data propose to resume the investigation with the RBF network. The RBF network can be also readily applied from the
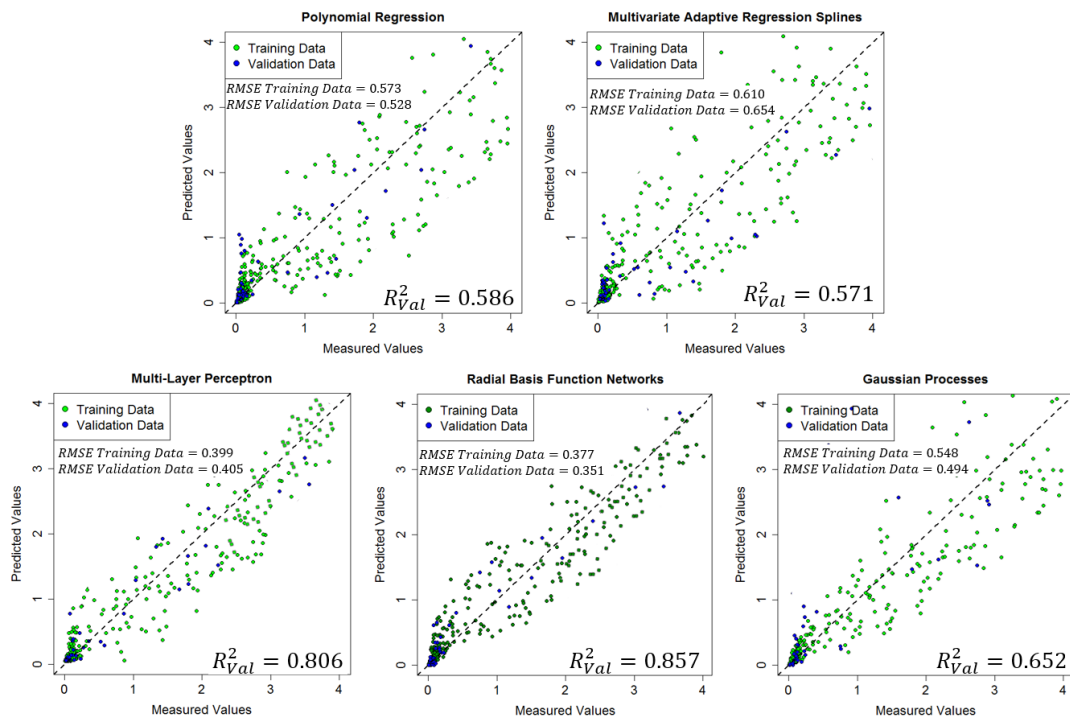
Figure 7.17: Metamodeling Approaches for Y_3, $N_{\mathrm{Sf}} = 500$

perspective of the simulation speed, as the stopping times are by and large conforming with results of table 7.4. With reference to section 6.5, a representative coverage of the two desired distributions (one for supplier A, the other for supplier B) is achieved by feeding the RBF network with $N' = 250.000$ pseudo MC samples of the underlying feature space distributions. Figures 7.18 and 7.19 show the resulting distributions for supplier A and B, respectively. There, the difference between the shapes of both distributions and the higher risk under supplier A are clearly visible. A sample of these distributions

corresponds indeed to a ISC testing result of a particular engine.
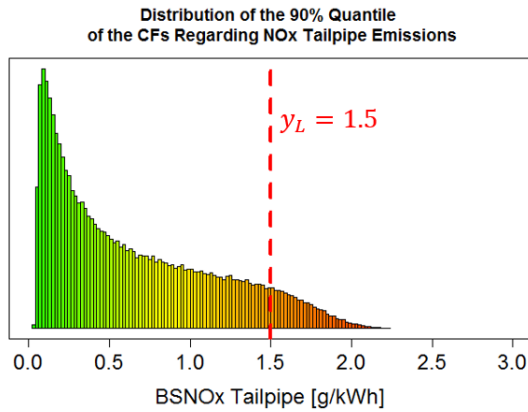
Figure 7.18: Distribution of Y_3 under Supplier A
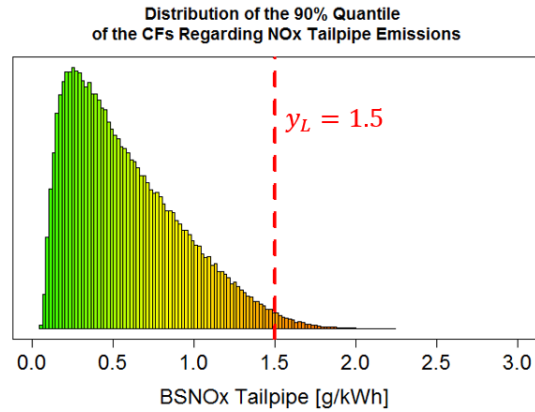
Figure 7.19: Distribution of Y_3 under Supplier B

The statistical sequential sampling plan, described in subsection A.2.4, is simulated for these distributions so that the associated failure risks and expected testing effort become eventually assessable for both NO_x sensors. Similarly to CoP, the simulation of the ISC testing procedure can be arbitrarily often repeated, such that probabilities for the number of engines to be tested and their according failure ratios become reasonably estimable. Figure 7.20 shows the resulting probabilities and failure fractions for both suppliers. Particularly, it becomes instantly clear that the ISC testing procedure can
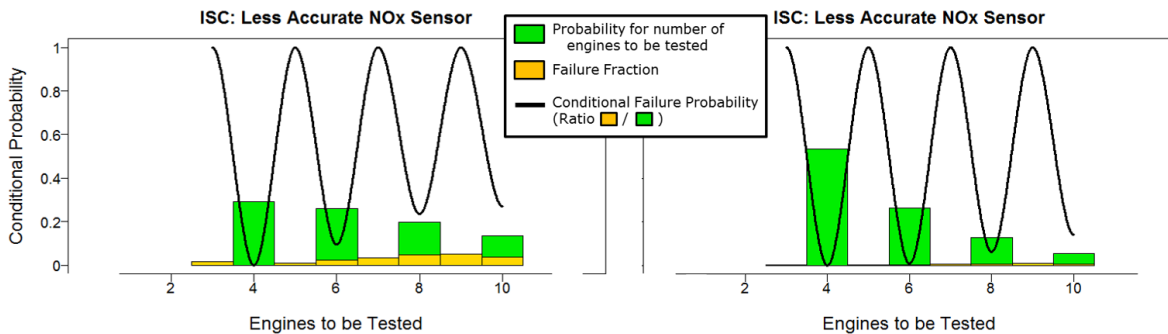
Figure 7.20: Risk Assessment ISC Testing Procedure, Left Supplier A, Right Supplier B

only be passed for an even number of tested engines. The sequential testing procedure is either continued with an additional engine or rejected when an odd number of engines is tested. In contrast to CoP, where absolutely no risk was present under supplier B, now for the ISC testing procedure there exists a positive failure probability for both suppliers. Nevertheless, supplier A entails a significant higher risk for a failure than supplier B. Moreover, the expected number of engines to be tested differs by more than 1 engine (supplier A: 6.48 engines, supplier B: 5.43 engines) for both strategies. Moreover, the overall failure probability can be decreased from 0.2213 down to 0.0383, if supplier B is

preferred over supplier A.

The case study presented in this chapter emphasizes how the production and lifetime-oriented development process can be employed in order to support an OEM in finding an optimal quality standard in terms of TCO. Thus, the new approach made it particularly possible to assess the risks during the CoP and ISC testing procedures for different quality standards (considered by two different accurate NO_x sensors). The final decision regarding the effectively applied quality standard, that is either supplier A or B, is usually taken by the OEM. It is eventually the vehicle manufacturer, who has to balance the acquired results with all associated costs.

Chapter 8

Conclusion and Outlook

The application of arbitrary technology is more restricted than ever. The globalization and the concurrent opening of the markets force production facilities to optimize relevant selling factors, like price level, product quality and included warranties. Nowadays, decisions to purchase are no longer influenced by the acquisition price, but are also dependent on the total costs of ownership (TCO), which embrace all costs incurred during the product's life cycle. Particularly for the automotive industry, the *production and lifetime-oriented development* was presented, which allows to optimally advise an OEM in minimizing the TCO under the constraining market requirements.

The quality standard defined by the OEM during the development stage determines for the most part the actual TCO. It is worth investing as much effort as technically feasible into that stage. OEMs frequently commission automotive consulting firms to receive an optimal support in defining their quality standards. This applies to the investigation of feasible hardware and software solutions, such as necessary EAS solutions, air path concepts or ECU control strategies. Particular attention is given the quality standards suggested to ensure that they comply with the overall requirements of all parts, the OEM, the public society and the end-users involved. Nevertheless, associated compliance risks are often not quantitatively measurable, but are exclusively subject to expert knowledge. Moreover, the variability of the production and usage behavior excessively aggravates the prediction of TCO for an implemented quality standard. Hidden costs arising from product failures, warranty services or legal penalties are considered as significant cost drivers that have to be avoided by all means. Solving the TCO minimization task requires addressing these open issues. In this regard, a new approach *the production and lifetime-oriented development* process was introduced in this thesis. The new approach proved to be capable of quantitatively assessing multiple quality standards with respect to given requirements and hidden costs. The basic idea is to mathematically incorporate the uncontrollable stochastic factors of the production and lifetime usage into the development process. On an experimental level, these can be jointly investigated with the controllable factors, which define the quality standards to be analyzed. Since MoBEO is slightly faster-than-real-time, the main difficulty lies in reaching a representative coverage of the stochastic factors with a manageable number of experiments. The basis of the new approach inevitably entails the appropriate modeling of the stochastic factors by probability distributions. In respect of the statistical process control, process capa-

bility is adapted and extended to other than normal distributions. The versatile Beta distribution and the non-parametric kernel approach prove to be suitable instruments, especially when the symmetric normal distribution is not applicable. Particularly, the analysis of usage profiles and their treatment by the multinomial distribution is valuable for modeling the stochastic usage behavior. The set of all feasible factor combinations, referred to as the *feature space*, often consists of more than 50 dimensions so that most efficient experimental designs are required to overcome the high numerical complexity. This doctoral thesis proposes to start the experimental procedure by a screening procedure, and more particularly, by the definitive screening design. This is because it provides "*nice to have*" characteristics in terms of first and second order effects using only twice as many experiments than factors considered. The subsequent investigation is then focused on the region of the feature space, which belongs to the most significant factors selected by the screening procedure. During that space-filling procedure the application of the Latin-Hypercube design is suggested for all factors modeling manufacturing tolerances. This is due to the fact that the associated distributions may depend on the quality standard investigated. The very efficient depth design, which had been introduced in this thesis, is suggested for all factors, whose associated distributions are considered to be definite. Thereby, the application of semi-physical aging models ensures the simulation of the generated experiments considering one or multiple system aging states. The employment of MC simulation procedures on "*up-to-date*" metamodel approaches turns out to be a powerful approach on how to achieve representative coverages for different feature space distributions with only one experimental run-through of MoBEO. With focus on decisive response variables, like emissions, fuel consumption or component lifetimes, the methodology established allows to investigate feasible quality standards under the stochastic factors of the production and lifetime usage. Eventually, the model-based simulation approach combined with statistical evaluation tools can significantly contribute to a better solution of the TCO minimization task.

The established methodology has so far achieved great interest among AVL customers, and the first commissioned projects have been successfully completed to full satisfaction. In order to further increase the market acceptance, several activities are scheduled for the near future. The influence of hardware tolerances on functional tolerances, as discussed in subsection 3.1.2, shall be examined in more detail. This especially concerns the impact of hardware tolerances of engine components on production costs, operating costs and hidden costs due to failure and OBD related issues. An additional objective is the initiation of a component supplier network so that hardware tolerances can be optimally modeled and better assessed in terms of TCO. A task in the near future will also be the investigation of tolerances in fuel and DEF quality. Closely linked to that is the intention to incorporate more realistic usage and aging behavior so that OBD events and warranty payments become better predictable. In that respect, the new depth design, introduced in chapter 4, is the perfect tool for efficiently modeling different usage profiles in a representative manner. Last but not least, the better understanding and incorporation of the ECU into the production and lifetime-oriented development process is another task that has to be mastered in the upcoming years.

Appendix A

Appendix

A.1 Vehicle Categories

All motor vehicles applicable to Directive 2007/46/EC are classified in the proper Annex II as follows:

Category M: Motor vehicles with at least four wheels, designed and constructed for the carriage of passengers

> **Category** M_1: Vehicles designed and constructed for the carriage of passengers and comprising no more than eight seats in addition to the driver's seat.
>
> **Category** M_2: Vehicles designed and constructed for the carriage of passengers, comprising more than eight seats in addition to the driver's seat, and having a maximum mass not exceeding 5 tonnes.
>
> **Category** M_3: Vehicles designed and constructed for the carriage of passengers, comprising more than eight seats in addition to the driver's seat, and having a maximum mass exceeding 5 tonnes.

Beside that, directive 2001/85/EC classifies vehicles having a maximal capacity of 22 passengers without the driver into class A and class B.

> **Class A**: Vehicles designed to carry standing passengers; a vehicle of this Class has seats and shall have provision for standing passengers.
>
> **Class B**: Vehicles not designed to carry standing passengers; a vehicle of this Class has no provision for standing passengers.

If 22 passengers without the driver are exceeded, directive 2001/85/EC differs three different classes, class I, II and III.

> **Class I**: Vehicles constructed with areas for standing passengers, to allow frequent passenger movement.
>
> **Class II**: Vehicles constructed principally for the carriage of seated passengers, and

designed to allow the carriage of standing passengers in the gangway and/or in an area which does not exceed the space provided for two double seats.

Class III: Vehicles constructed exclusively for the carriage of seated passengers.

Category N**:** Motor vehicles with at least four wheels designed and constructed for the carriage of goods.

Category N_1: Vehicles designed and constructed for the carriage of goods and having a maximum mass not exceeding 3,5 tonnes.

Category N_2: Vehicles designed and constructed for the carriage of goods and having a maximum mass exceeding 3,5 tonnes but not exceeding 12 tonnes.

Category N_3: Vehicles designed and constructed for the carriage of goods and having a maximum mass exceeding 12 tonnes.

Additionally, Directive 2007/46/EC distinguishes between on-road and off-road vehicles (symbol G), which are defined in the following way:

Category G**:** All vehicles of class N_1 with a maximum mass not exceeding 2.000 kg and of M_1 are considered as off-road vehicles, if

- at least one front axle and at least one rear axle is designed to be driven simultaneously, including vehicles where the drive to one axle can be disengaged,

- at least one differential locking mechanism or at least one mechanism having a similar effect is obstructed, and if the vehicle can climb a 30% gradient,

- and at least five of the following six requirements have to be satisfied:

    - the approach angle must be at least 25 degrees

    - the departure angle must be at least 20 degrees

    - the ramp angle must be at least 20 degrees

    - the ground clearance under the front axle must be at least 180 mm

    - the ground clearance under the rear axle must be at least 180 mm

    - the ground clearance between the axles must be at least 200 mm.

Vehicles of class N_1 with a maximum mass exceeding 2.000 kg and vehicles of classes N_2, M_2 and M_3 with a maximum mass not exceeding 12.000 kg are considered as off-road vehicles, if

- all their wheels are designed to be driven simultaneously, including vehicles where the drive to one axle can be disengaged, or if

- the following three points apply

    - at least one front and at least one rear axle are designed to be driven simultaneously, including vehicles where the drive to one axle can be disengaged

- there is at least one differential locking mechanism or at least one mechanism having a similar effect

- they can climb a 25% gradient calculated for a solo vehicle.

Vehicles of class M_3 with a maximum mass exceeding 12.000 kg and vehicles of classe N_3 are denoted as off-road vehicles if

- all their wheels are designed to be driven simultaneously, including the cases, where the drive to one axle can be disengaged

- or if the following requirements are met

  - at least half the wheels are driven

  - there is at least one differential locking mechanism or at least one mechanism having a similar effect

  - they can climb a 25% gradient calculated for a solo vehicle,

  - at least four of the following six requirements are satisfied:

    · the approach angle must be at least 25 degrees,
    · the departure angle must be at least 25 degrees,
    · the ramp angle must be at least 25 degrees,
    · the ground clearance under the front axle must be at least 250 mm,
    · the ground clearance between the axles must be at least 300 mm,
    · the ground clearance under the rear axle must be at least 250 mm.

All vehicles, which cannot be classified as an off-road vehicle, are said to be 'on-road' vehicles and are, in principle, only intended for the use on the road.

A.2   European Regulatory Framework and the Emission Standard EURO VI

The European basic legislative framework of motor vehicles, their trailers and of all kind of equipment designed for such vehicles is rooted in Directive 70/156/**E**uropean **E**conomic **C**ommunity (EEC), established by the EEC in 1970. Until 2007, Directive 70/156/EEC has been amended several times so that the EC established Directive 2007/46/EC for reasons of transparency. The directive regulates all administrative provisions and general technical requirements, relevant for the approval of new vehicles and all of their components with the exception of agricultural equipment or forestry tractors, as defined in Directive 2003/37/EC, quadricycles, as defined in Directive 2002/24/EC, and tracked vehicles. In the European Union emission requirements are primarily dependent on the vehicle reference mass. Vehicles of classes M_1, M_2, N_1 and N_2 with a reference mass[1] not exceeding 2610 kg are denoted as *light-duty* applications. The class of *heavy-duty* applications consists of all vehicles of categories M_1, M_2, N_1 and N_2 with a reference mass

---

[1]Reference mass is defined as the mass of the vehicle in running order +25 kg

exceeding 2610 kg and all vehicles of classes M_3 and N_3.

The approval of a new vehicle or engine type normally requires tests with the complete vehicle. However, in case of heavy-duty applications it usually suffices to test the engine environment of the vehicle. The test object, as described in section 7.1, and all other on-road heavy-duty applications powered by CI engines, **P**ositive **I**gnition (PI) natural gas or liquefied petroleum gas engines are subject to emission regulations denoted by the synonym *Euro VI*, as shortly described in the following subsections.

A.2.1 Emission Regulation

The original basis of European emission regulations of all vehicles exceeding a maximum design speed of $25 \, km/h$ is Directive 88/77/EEC of 3rd December 1987. The same was later re-cast and consolidated in the Directive 05/55/EC of 2005, where durability and OBD specifications were introduced and emission limits were re-defined. A clear legislative framework for motor vehicles and their trailers and equipment was established in September 2007, as provided by Directive 2007/46/EC ("*framework directive*"). Manifested by Regulation (EC) No 595/2009 and amended by the Commission Regulation (EU) No 582/2011, the EURO VI norm introduced more restrictive emission limits. A limit for the **P**article **N**umber (PN), stricter OBD requirements and new testing requirements like off-cycle and ISC testing were added.[2] Commission Regulation (EU) No 582/2011 particularly refers to Regulation (EC) No 49 ECE of the **E**conomic **C**ommission for **E**urope *of the **U**nited **N**ations* (UN/ECE) in terms of all measures to be undertaken against the emission of gaseous and particulate pollutants from CI engines and from PI engines powered by natural gas or liquefied petroleum gas. Beside the European emission standards for on-road heavy-duty applications, different emission standards are existent worldwide. Examples are the *United States-Standard 2010* or the *Japanese **P**ost-**N**ew **L**ong **T**erm **R**egulations* (PNLTR). Table A.1 provides an overview of the change in legislation over the last years and with regard to the future for these regions. With regard to the EURO VI norm, Article 12(2) of Regulation (EC) No 595/2009 dictates emission test procedures for the approval of a new vehicle or engine type. The main test procedures to be conducted for a heavy-duty on-road application are the **W**orld **H**armonized **S**tationary **C**ycle (WHSC) and the **W**orld **H**armonized **T**ransient **C**ycle (WHTC), both defined in Annex 4 of Regulation (EC) No 49 ECE.

The WHSC, which is provided in table A.2, is a steady-state emission test that is characterized by a schedule of engine operation points, consisting of speed and torque. The operation points have to be maintained for predetermined *time spans*, wherein recordings are made every 0.1 s (10 Hertz). During the WHSC the tested system has time to stabilize in the sense that mass flow values, temperatures and emissions maintain almost constant levels. The operation points are processed linearly within 20 ± 1 seconds. The purpose of the weighting factors entailed is to highlight the portion, spent on the single operation points. Before starting a WHSC test, the engine has to be warmed up on operation point "9" for at least 10 minutes, but no longer than 30 minutes, and then needs to be turned off for 5 minutes. Operation point "0" is not executed but considered with zero emissions

---

[2] https://www.dieselnet.com

US Emissions (g/kWh)	EU Emissions (g/kWh, g/kWh)	Japan Emissions (g/kWh)

ESC / HDDTC — US 2007
- NOx 1,58
- PM 0,0134
- CO 20,79
- HC 0,19

RMC / HDDTC — US 2010
- NOx 0,268
- PM 0,0134
- CO 20,79
- HC 0,19

ESC / ETC — Euro IV
- NOx 3,5
- PM 0,02 / 0,03
- CO 1,5 / 4,0
- HC 0,46 / 0,55   NH3 25ppm

ESC / ETC — Euro V
- NOx 2,0
- PM 0,02 / 0,03
- CO 1,5 / 4,0
- HC 0,46 / 0,55
- NH3 25ppm

WHSC / WHTC — Euro VI
- NOx 0,4 / 0,46
- PM 0,01
- PN 8 x 10^{11} / 6 x 10^{11}
- CO 1,5 / 4,0
- HC 0,13 / 0,16
- NH3 10 ppm

JE05 — NLTR
- NOx 2,0
- PM 0,027
- CO 2,22
- HC 0,17

JE 05 — PNLTR
- NOx 0,7
- PM 0,01
- CO 2,22
- HC 0,17

WHTC / Fuel Efficiency (JE05) — 2016 PPNLTR
- NOx 0,40
- PM 0,01
- CO 2,22
- HC 0,17

Table A.1: Emissions Limits of On-Road Heavy-Duty Applications

and zero power. It is supposed to influence the final result with a weight of 24%.[3] The relevant key figures are the brake specific emission values. These are determined by the fraction of the individual emission mass flow (m_{gas} in $[mg]$) and integrated force (P in $[kW]$) over the testing procedure (this is the work W_{test}), calculated by

$$e_{\mathrm{gas,test}} = \frac{\int_{Test} m_{\mathrm{gas}}}{\int_{Test} P} \approx \frac{\sum_t m_{\mathrm{gas}}(t) \cdot 0.1}{\sum_t P(t) \cdot 0.1}. \tag{A.1}$$

The WHTC, depicted in figure A.1[4], is a 30 minutes long-lasting series of speed and torque conditions for heavy-duty on-road engines, each run for 0.1 s (10 Hertz). Despite the WHSC, the short maintenance at each operation point forces the tested system to operate in a dynamic manner. This means that the tested system does not necessarily have enough time to stabilize. The idea behind the transient concept is to simulate real driving behavior of vehicle, as discussed in the form of usage profiles in subsection 3.2.2. This includes cold and hot starting, frequent accelerations, changes of load, idling, urban driving, rural driving and highway driving. The cold and the hot WHTC test procedures (WHTC_cold and WHTC_warm) are considered separately. After completion of the cold start test, the powertrain is conditioned with a 10 ± 1 minutes long-lasting soak period for the hot start test. A weighted average of the cold start emission result (14%) and the hot start emission result (86%) determines the final test result. It is given by

$$e_{\mathrm{gas,WHTC}} = 0.14 \cdot e_{\mathrm{gas,WHTC\_cold}} + 0.86 \cdot e_{\mathrm{gas,WHTC\_warm}} \tag{A.2}$$

---

[3]cf. www.dieselnet.com
[4]image source: transportpolicy.net

Operation Point	Speed	Load	Time Span	Weighting Factor
–	%	%	s	
0	0	–	–	0.24
1	0	0	210	0.085
2	55	100	50	0.02
3	55	25	250	0.10
4	55	70	75	0.03
5	35	100	50	0.02
6	25	25	200	0.08
7	45	70	75	0.03
8	45	25	150	0.06
9	55	50	125	0.05
10	75	100	50	0.02
11	35	50	200	0.08
12	35	25	250	0.10
13	0	0	210	0.085
Total			**1895**	**1**

Table A.2: World Harmonized Stationary Cycle (WHSC)

The WHSC and WHTC have to be conducted under the *normal conditions of use.* Normal conditions of use denote an atmospheric pressure (pb) that is greater or equal 82.5 kPa and an ambient temperature T_{amb} that needs to exceed at least 266 K (or $-7\,°$C) but that is not allowed to exceed the boundary $-0.4514\cdot(101.3-pb[kPa])+311 \overset{!}{>} T_{\mathrm{amb}}$. In addition to that, the engine coolant temperature $T_{\mathrm{coolant}} \overset{!}{>} 343$ K (or $T_{\mathrm{coolant}} \overset{!}{>} 70\,°$C).

A.2.2   Emission Durability

In addition to Article 12(2) of Regulation (EC) No 595/2009, where the emission compliance of a new vehicle or engine has to be tested, Article 3 of Directive 05/55/EC requires the compliance with the emission limits during the *useful life* of a vehicle or engine to be approved (cf. table A.3). The useful life depends on the vehicle category. In order to com-

Vehicle Category	Durability (km or years, whichever is sooner)	
M$_1$, N$_1$ and M$_2$	160 000 km	5 years
N$_2$, N$_3$ with a maximum technically permissible mass not exceeding 16 tonnes and M$_3$ Class I, Class II and Class A, and Class B with a maximum technically permissible mass not exceeding 7.5 tonnes	300 000 km	6 years
N$_3$ with a maximum technically permissible mass exceeding 16 tonnes and M$_3$, Class III and Class B with a maximum technically permissible mass exceeding 7.5 tonnes.	700 000 km	7 years

Table A.3: Emission Durability

ply with the emission limits at the end of useful life, aging behavior has to be respected in terms of emission deterioration. Exact deterioration forecasts are rarely available for
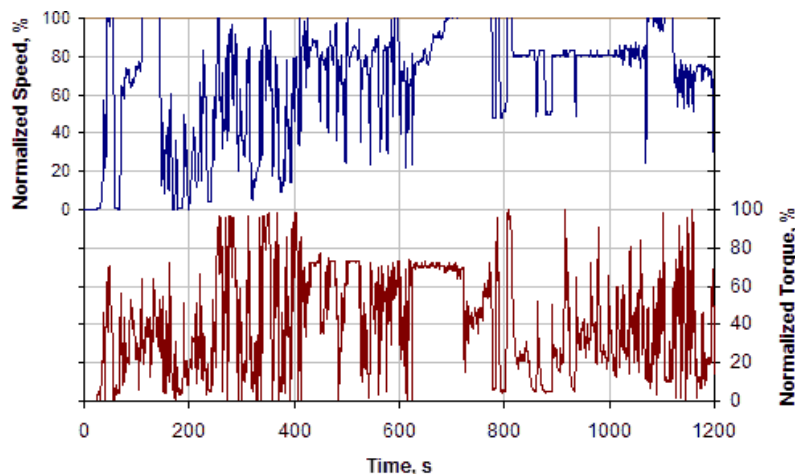
Figure A.1: World Harmonized Transient Cycle (WHTC)

the manufacturer at the time of the engine approval. Therefore, Regulation (EC) No 595/2009 provides legislative **D**eterioration **F**actors (DFs), as outlined by table A.4. DFs may also be estimated by a predetermined service accumulation schedule that includes at least three hot WHSC and/or the WHTC test runs. In any case, two test runs have to be set at the beginning and at the end of the service schedule. The service accumula-

Deterioration factors

Test cycle	CO	THC ([1])	NMHC ([2])	CH_4 ([2])	NO_x	NH_3	PM mass	PM number
WHTC	1,3	1,3	1,4	1,4	1,15	1,0	1,05	1,0
WHSC	1,3	1,3	1,4	1,4	1,15	1,0	1,05	1,0

([1]) Applies in case of a compression-ignition engine.
([2]) Applies in case of a positive-ignition engine.

Table A.4: Assigned Deterioration Factors

tion schedule itself may end before useful life is reached, but it must not be shorter than outlined in table 3.2.1.8. of Regulation (EC) No 582/2011. Emission results, which are recorded during the service accumulation schedule, have to be considered with one plus the same number of decimal digits and shall be fitted with a linear regression model (cf. table A.2), as specified in Annex I to Regulation (EC) No 595/2009. An optional prior agreement of the approval authority may permit the application of non-linear regression approaches. Emission values of all pollutants shall be calculated at the beginning and at the end of useful life, if necessary by extrapolating the regression equation. Although individual test points may exceed the emission limits, the emissions calculated at these points have to meet the emission limits at all costs (cf. table A.5). The multiplicative DF results from the fraction of the calculated emission result at the end of useful life and the emission result at the beginning of the service accumulation schedule. If agreed with the
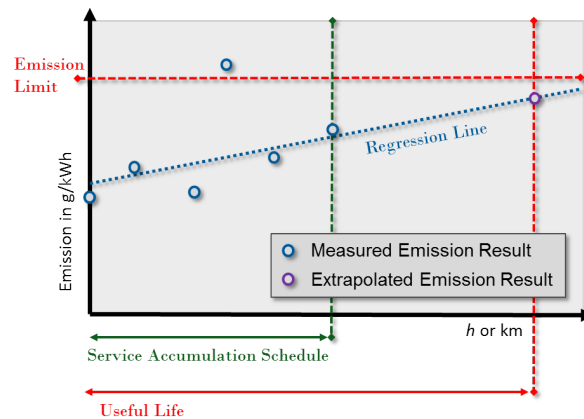
Figure A.2: Determination of the Deterioration Factor

Euro VI Emission Limits

	Limit values							
	CO (mg/kWh)	THC (mg/kWh)	NMHC (mg/kWh)	CH$_4$ (mg/kWh)	NO$_x$ (^{1}) (mg/kWh)	NH$_3$ (ppm)	PM mass (mg/kWh)	PM (^{2}) number (#/kWh)
WHSC (CI)	1 500	130			400	10	10	8.0×10^{11}
WHTC (CI)	4 000	160			460	10	10	6.0×10^{11}
WHTC (PI)	4 000		160	500	460	10	10	(^{3})

PI = Positive Ignition.
CI = Compression Ignition.
(^{1}) The admissible level of NO$_2$ component in the NO$_x$ limit value may be defined at a later stage.
(^{2}) A new measurement procedure shall be introduced before 31 December 2012.
(^{3}) A particle number limit shall be introduced before 31 December 2012.'

Table A.5: Emission Limits - EURO VI

approval authority, the use of an additive DF is possible. This factor corresponds to the difference of the emission results, obtained at the end and at the start of the service accumulation schedule. If the manufacturer makes the decision to estimate the DF for only one of the test cycles (WHSC or WHTC), emission results of the other emission cycles should be calculated at the beginning and the end of the service accumulation schedule. These have to remain applicable for the calculated DFs of each pollutant.

A.2.3 Conformity of Production Requirements

With regard to paragraph 8 of Regulation (EC) No 49 ECE CoP shall be tested with a minimum sample size of three vehicles or engines by one of the three tests, presented subsequently. An acceptance sampling system for inspection by variables (these are measuring methods) shall be applied for the first two test procedures. The first test requires a known and satisfactory production standard deviation for each pollutant to be tested.

The approval authority considers the standard deviation of a pollutant as satisfactory, if respective emission measurement data adheres arranged limits. These limits are usually set by a standard or a contract in the forefront of CoP (cf. ISO 3951-5:2006). In case the pollutant's standard deviation is either unknown or unsatisfactory, the second test procedure, the so called *s*-method needs to be applied. At manufacturer's request the third test can be carried out. It specifies an acceptance sampling system for inspection by attributes (these are counting methods) that counts the number of vehicles or engines that exceed the legal emission limits (cf. ISO 8422:1991).

Test 1: σ-Method

The so called σ-method is applied for CoP testing, if the production standard deviation is satisfactory for the approval authority. Appendix 1 of Regulation (EC) No 49 ECE describes this method in the following way.

With a minimum sample size of three engines the sampling procedure is set so that the probability of a lot passing a test with 40 per cent of the engines defective is 0.95 (producer's risk = 5 per cent), while the probability of a lot being accepted with 65 per cent of the engines defective is 0.10 (consumer's risk = 10 per cent).

The following test statistic has to be evaluated

$$T_N = \frac{1}{\sigma} \sum_{n=1}^{N} (L - y_n) = N \cdot \frac{L - \bar{y}}{\sigma}, \tag{A.3}$$

where

$\quad L$ *... the natural logarithm[5] of the limit value for the pollutant,*

$\quad y_n$ *... the natural logarithm of the measurement (after having applied the relevant DF) for the n-th engine or sample,*

$\quad \sigma$ *... the production standard deviation of the tested pollutant,*

$\quad N$ *... the current sample number.*

Then:

(a) *If the test statistic result is greater than the pass decision number for the sample size given in table A.6, a pass decision is reached for the tested pollutant;*

(b) *If the test statistic result is less than the fail decision number for the sample size given in table A.6, a fail decision is reached for the tested pollutant;*

(c) *Otherwise, an additional engine is tested in accordance to paragraph 8.3.2. (of Regulation (EC) No 49 ECE).*

The pass and fail decision numbers, as outlined in table A.6, are subject to ISO 3951:2013.

---

[5]ISO assumes pollutants to be Lognormally distributed random variables

Test 2: s-Method

If the production standard deviation is either unknown or unsatisfactory, the so called s-method shall be applied. The respective test procedure is provided by appendix 2 of Regulation (EC) No 49 ECE as follows.

With a minimum sample size of three engines the sampling procedure is set so that the probability of a lot passing a test with 40 per cent of the engines defective is 0.95 (producer's risk = 5 per cent), while the probability of a lot being accepted 65 per cent of the engines defective is 0.10 (consumer's risk = 10 per cent).

Thereby, the subsequent test statistic has to be evaluated

$$T_N = \frac{1/N \cdot \sum_{n=1}^{N} (y_n - L)}{\sqrt{1/N \cdot \sum_{n=1}^{N} (y_n - \bar{y})^2}} = \frac{\bar{y} - L}{s \cdot \sqrt{\frac{N-1}{N}}}. \tag{A.4}$$

The test statistic determines whether the production series has passed or failed the test:

(a) *Pass the test, if $T_N \leq A_N$;*

(b) *Fail the test, if $T_N \geq B_N$;*

(c) *Draw another engine, if $A_N < T_N < B_N$.*

Table A.7 contains the values of the pass decision numbers (A_N), the fail decision numbers (B_N) and the corresponding sample sizes. Similarly to the σ-method, the pass and fail decision numbers follow from ISO 3951:2013.

Test 3: Inspection by Attributes

In coordination with the approval authority an acceptance sampling system for inspection by attributes, as outlined by appendix 3 of Regulation (EC) No 49 ECE, can be carried out. This procedure reads in the following way.

With a minimum sample size of three engines the sampling procedure is set so that the probability of a lot passing a test with 30 per cent of the engines defective is 0.90 (producer's risk = 10 per cent), while the probability of a lot being accepted 65 per cent of the engines defective is 0.10 (consumer's risk = 10 per cent).

(a) *If the test statistic is less than or equal to the pass decision number for the sample size given in table A.8, a pass decision is reached for the pollutant;*

(b) *If the test statistic is greater than or equal to the fail decision number for the sample size given in table A.8, a fail decision is reached for the pollutant;*

(c) *Otherwise, an additional engine is tested according to paragraph 8.3.2. of this regulation and the calculation procedure is applied to the sample increased by one more unit.*

The pass and fail decision numbers, as provided by table A.8, refer to ISO 8422:1991.

Cumulative number of engines tested (sample size)	Pass decision number A_n	Fail decision number B_n
3	3.327	− 4.724
4	3.261	− 4.790
5	3.195	− 4.856
6	3.129	− 4.922
7	3.063	− 4.988
8	2.997	− 5.054
9	2.931	− 5.120
10	2.865	− 5.185
11	2.799	− 5.251
12	2.733	− 5.317
13	2.667	− 5.383
14	2.601	− 5.449
15	2.535	− 5.515
16	2.469	− 5.581
17	2.403	− 5.647
18	2.337	− 5.713
19	2.271	− 5.779
20	2.205	− 5.845
21	2.139	− 5.911
22	2.073	− 5.977
23	2.007	− 6.043
24	1.941	− 6.109
25	1.875	− 6.175
26	1.809	− 6.241
27	1.743	− 6.307
28	1.677	− 6.373
29	1.611	− 6.439
30	1.545	− 6.505
31	1.479	− 6.571
32	− 2.112	− 2.112

Table A.6: σ-Method: Pass and Fail Decision Numbers

Cumulative number of engines tested (sample size)	Pass decision number A_n	Fail decision number B_n
3	-0.80381	16.64743
4	-0.76339	7.68627
5	-0.72982	4.67136
6	-0.69962	3.25573
7	-0.67129	2.45431
8	-0.64406	1.94369
9	-0.61750	1.59105
10	-0.59135	1.33295
11	-0.56542	1.13566
12	-0.53960	0.97970
13	-0.51379	0.85307
14	-0.48791	0.74801
15	-0.46191	0.65928
16	-0.43573	0.58321
17	-0.40933	0.51718
18	-0.38266	0.45922
19	-0.35570	0.40788
20	-0.32840	0.36203
21	-0.30072	0.32078
22	-0.27263	0.28343
23	-0.24410	0.24943
24	-0.21509	0.21831
25	-0.18557	0.18970
26	-0.15550	0.16328
27	-0.12483	0.13880
28	-0.09354	0.11603
29	-0.06159	0.09480
30	-0.02892	0.07493
31	-0.00449	0.05629
32	0.03876	0.03876

Table A.7: s-Method: Pass and Fail Decision Numbers

Cumulative number of engines tested (sample size)	Pass decision number	Fail decision number
3	—	3
4	0	4
5	0	4
6	1	5
7	1	5
8	2	6
9	2	6
10	3	7
11	3	7
12	4	8
13	4	8
14	5	9
15	5	9
16	6	10
17	6	10
18	7	11
19	8	9

Table A.8: CoP at Manufacturer's Request: Pass and Fail Decision Numbers

A.2.4 In-Service Conformity Requirements

A schedule and a sampling plan must be provided for ISC by the manufacturer within 18 months from the first registration of type approved vehicle. ISC testing shall be repeated every two years over the useful life period of the tested vehicle or engine type. Annex 8 of Regulation (EC) No 49 ECE sets out a statistical sequential sampling plan for an inspection by attributes (these are counting methods) in order to test the ISC requirements (cf. ISO 8422:1991). The corresponding test procedure reads as follows.

With a minimum sample size of three engines the sampling procedure shall be set so that the probability of a lot passing a test with 20% of the vehicles or engines defective is 90% (producer's risk = 10%) while the probability of a lot being accepted with 60% of the vehicles or engines defective is 10%. The test statistic to be evaluated is defined by the number of non-conforming vehicles or engines given a lot size that equals N. The pass or fail decision of the lot shall be made according to the following requirements:

(a) If the test statistic is less than or equal to the pass decision number for the sample size given in table A.9, a pass decision is reached for the lot;

(b) If the test statistic is greater than or equal to the fail decision number for the sample size given table A.9, a fail decision is reached for the lot;

(c) Otherwise, an additional engine is tested, and the calculation procedure is applied to the sample increased by one more unit.

Cumulative Number of Engines Tested (Sample Size)	Pass Decision Number	Fail Decision Number
3	-	3
4	0	4
5	0	4
6	1	4
7	1	4
8	2	4
9	2	4
10	3	4

Table A.9: In-Service Conformity: Pass and Fail Decision Numbers

Whether a test vehicle or engine is conforming or not depends on the **C**onformity **F**actors (CFs), which have to be determined by the use of the *work based averaging window method* for all regulated pollutants. There, the idea is to evaluate measurement data, recorded by PEMS for sequential sub-sets or so called *windows*. The lengths of these windows are determined as to match the work of the WHTC (W_{WHTC}). As depicted in figure A.3, the length of the i-th window $t_i = [t_{1,i}, t_{1,i}]$ is chosen in a way that $W(t_{2,i}) - W(t_{1,i}) \geq W_{\mathrm{ref}}$ and $W(t_{2,i} - \Delta t) - W(t_{1,i}) < W_{\mathrm{ref}}$. The time increment Δt of the moving average calculations to be performed corresponds to the data sampling period. It should be chosen less or equal than 1 second (i.e. $\Delta t \leq 1s$). All windows with average power
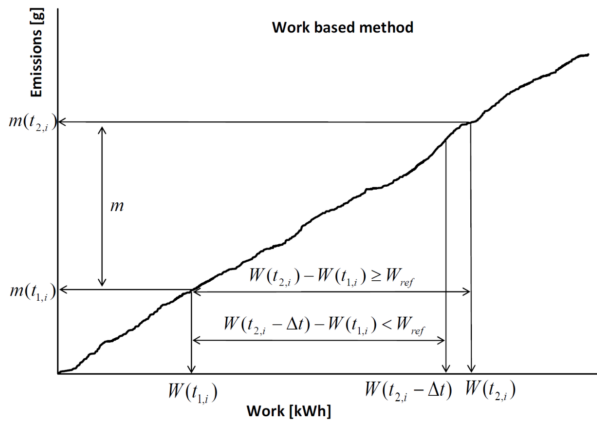
Figure A.3: Work Based Method

Pollutant of Engines Tested	Maximum allowed CF
CO	1.50
THC	1.50
NMHC	1.50
CH_4	1.50
NO_x	1.50

Table A.10: In-Service Conformity: Maximum Allowed Conformity Factors

less than 20% of the maximum engine power are invalid windows and must be neglected. Moreover, the ratio of valid windows has to be equal or greater than 50%. Therefore, lower power thresholds can be applied, if the 50% boundary is not reached. In such cases, the power threshold has to be successively decreased by 1%, but not below 15%. For the i-th window the specific emission e_{gas} (mg/kWh) is calculated by the fraction of mass

emissions and work

$$e_{\text{gas},t_i} = \frac{\int_{[t_{1,i},t_{2,i}]} m_{\text{gas}}}{W\left(t_{2,i}\right) - W\left(t_{1,i}\right)} \approx \frac{\sum_{t \in [t_{1,i},t_{2,i}]} m_{\text{gas}}(t)}{\sum_{t \in [t_{1,i},t_{2,i}]} P(t)\Delta(t)}. \tag{A.5}$$

CFs shall be calculated for all valid windows and all relevant pollutants by

$$CF_{gas,i} = \frac{e_{\text{gas},t_i}}{L_{\text{gas}}}, \tag{A.6}$$

where L_{gas} is the corresponding emission limit of the WHTC, as specified by table A.5. With regard to the described test procedure, a test vehicle or engine is denoted to be non-conforming, if the 90% quantile of the calculated CFs turns out to be larger than the values outlined in table A.10.

A.3 Abbreviations, Acronyms and Symbols

α	parameter of the Beta distribution, confidence level
$\dot{\alpha}$	port flow coefficient (gas mass flow)
α_p	depth value enclosing a probability of p
β	parameter of the Beta distribution
χ	characteristic function, chi-squared distribution
$\boldsymbol{\varkappa}^t$	parameter vector, implementing an aged status at time t
Δ	random offset error of a sensor
Δ_n	record resolution of simulated experiment n
δ	realized offset error after calibration
ϵ	random model error (univariate regression)
$\boldsymbol{\Phi}$	design matrix (univariate regression)
ϕ	row vectors of $\boldsymbol{\Phi}$
φ	number of input factors for an aging model
Γ	random gain error
γ	gain error determined by sensor calibration, Gamma distribution
η	effectiveness, efficiency
ι	index, reserved for subunits of a technical unit
$\mathcal{K}(\cdot,\cdot)$	kernel function (univariate regression)
κ	parameter of the Gamma distribution
$\lambda(\cdot)$	Lebesgue measure
μ	parameter of the normal distribution, mean
$\boldsymbol{\mu}$	parameter vector of the multivariate normal distribution, mean
$\boldsymbol{\theta}$	parameter vector of an arbitrary distribution
θ	parameter of the Gamma distribution
$\boldsymbol{\Sigma}$	parameter of the multivariate normal distribution, variance matrix
σ	standard deviation, standard deviation of a production process
σ^2	parameter of the normal distribution, variance of a production process
τ	transfer function of sensor
Ω_j	domain of random variable X_j
Ξ	features representing the usage behavior
$\boldsymbol{\xi}$	multiple realizations of Ξ
ξ	realization of Ξ
$\boldsymbol{\Psi}$	observation matrix of $\boldsymbol{\xi}$

$\nabla_{\mathbf{a}}$	gradient of vector $\mathbf{a}$
$A\left(\cdot\,\middle|\,\mathbf{p}_A\right)$	aging model under parameterization $\mathbf{p}_A$
A_L	linear effect of factor A
A_l	circle with radius r_l
A_Q	quadratic effect of factor A
$AB_{L\times L}$	linear interaction effect of factors A and B
a	parameter of the Weibull distribution
a_j	activation of a neuron j
$\mathbb{B}_1^k$	k-dimensional ball with radius 1
B_r	multidimensional basis functions (univariate regression)
b	parameter of the Weibull distribution
C_{A_L}	linear contrast of factor A
C_{A_Q}	quadratic contrast of factor A
$C_{AB_{L\times L}}$	linear interaction contrast of factors A and B
C_{N_p}	general capability process value
C_p	capability process value
$C_{p\kappa}$	capability process value for production drifts
c	parameter of the Weibull distribution, coefficient vector (contrasts)
cdf	cumulative distribution function
$\mathbf{COV}(X_i, X_j)$	covariance between random variables X_i and X_j
$D(\cdot, F)$	depth function of distribution F
$D_{\boldsymbol{\varkappa}^t}$	damage status at point in time t implemented by $\boldsymbol{\varkappa}^t$
$DD_F(n)$	the n-th test run of a depth design for distribution F
d	order of a polynomial regression model
$df\mathbf{s}$	degrees of freedom
E	activation energy (Arrhenius equation)
$\mathbb{E}(X)$	expected value of random variable X
$e_{\mathbf{gas}}$	specific emission
edf	empirical distribution function
F	continuous probability distribution
$\mathcal{F}$	class of distributions on the *Borel sets*
G	graph (graph theory)
H	halfspace, Hessian matrix
$h(\cdot)$	activation function (univariate regression)

I	number of subunits per technical unit
$\mathbb{I}^k$	k-dimensional identity matrix
$\breve{i}$	index, reserved for robust features of the usage stage
$\mathbf{J}$	Jacobian matrix
j	index, reserved for the overall considered factors
$\check{j}$	index, reserved for manufacturing tolerances
$\tilde{j}$	index, reserved for open system parameters
$\breve{j}$	index, reserved for features of the usage stage
j'	index, reserved for aging models of engine components
j''	index, reserved for aging models of EAS components
K	rate constant (Arrhenius equation)
$\mathbf{K}$	Gram matrix (univariate regression)
k	total number of considered factors
$\check{k}$	number of features where manufacturing tolerances are considered
$\tilde{k}$	number of considered open system parameters
$\breve{k}$	number of features considered for usage space
k'	number of considered aging models for engine components
k''	number of considered aging models for EAS components
k_0	preexponential constant (Arrhenius equation)
$k_{\mathcal{R}}$	length of vector $\mathbf{w}$ (univariate regression)
k_s	reduced number of considered factors in terms of response y_s
L	random lifetime of a technical unit
$L_{\mathbf{gas}}$	legal emission limit
$L_2 D(\cdot, F)$	L_2-depth function of distribution F
$\tilde{L}_2 D(\cdot, F)$	affine invariant L_2-depth function of distribution F
l	realization of a random variable L
M	nominal value of a production process
M_F	multi-dimensional median of distribution F
$m_{\check{j}}$	design value of a feature, where manufacturing tolerances are considered
$\breve{m}$	number of robust features considered for usage space
N	number of total experiments that can be simulated, $N = N_{\mathrm{Sc}} + N_{\mathrm{Sf}}$
$\check{N}$	number of features, where manufacturing tolerances are considered
$\breve{N}$	number of real driving measurement observations
$\grave{N}$	number of observed lifetimes

$N_{\mathbf{Sc}}$	number of feasible screening experiments, $N = N_{\mathrm{Sc}} + N_{\mathrm{Sf}}$
$N_{\mathbf{Sf}}$	number of feasible space filling experiments, $N = N_{\mathrm{Sc}} + N_{\mathrm{Sf}}$
n	index, reserved for numbers
$O(\cdot, F)$	outlyingness function of a distribution F
P	usage profile, power [kW]
$\mathbb{P}(X \leq x)$	probability for a random variable X to realize smaller or equal than x
$\mathbf{p}_A$	fit parameter vector of aging model
$p_{\mathbf{emp}}$	empirical probability
p_l	probability associated to A_l
pdf	probability density function
$Q(\cdot, F)$	median-oriented quantile function for distribution F
q_p	theoretical quantile for probability p
R	universal gas constant (Arrhenius equation)
$\mathbb{R}$	set of the real numbers
$\mathbb{R}_+$	set of the real positive numbers
$\mathcal{R}$	univariate regression model
$\hat{\mathcal{R}}$	prediction of a univariate regression model
R^2	coefficient of determination
$R^2_{\mathbf{adj}}$	adjusted coefficient of determination
$R^2_{\mathbf{Val}}$	coefficient of determination aiming validation data
$R(\cdot, F)$	rank function of a distribution F
r	number of considered response variables
r_{ij}	Pearson correlation coefficient of random variables X_i and X_j
r_l	radius of circle A_l
r_n	n-th residuum
S	stimulus of a sensor
SS_{A_L}	linear effect square sum of factor A
SS_{A_Q}	quadratic effect square sum of factor A
$SS_{AB_{L \times L}}$	linear interaction effect square sum of factors A and B
ST_A	total sensitivity index of factor A
SSR	error sum of squares
SST	total sum of squares
s	index, reserved for response variables
s_i	empirical standard deviation of a random variable X_i

s_{ij}	empirical covariance of random variables X_i and X_j
T	tolerance interval (SPC), absolute temperature
t	time
$\breve{t}$	number of observations per feature $V_{\breve{j}}$ of the usage stage
U	uniformly distributed random variable (open system parameters)
u	realization of random variable U
V	feature regarding environmental conditions
$\mathbf{V}$	random vector modeling environmental conditions
W	physical work
$\mathbf{w}$	parameter vector (univariate regression)
$\hat{\mathbf{w}}$	estimated parameter vector (univariate regression)
$\mathbf{VAR}(X)$	variance of random variable X
$\mathbb{X}^k$	k-dimensional feature space
$\mathbf{X}$	random vector
X	random variable modelling stochastic factors
$\mathbf{x}$	realization of random vector $\mathbf{X}$
x	realization of a random variable X
$\mathbf{x}_*$	new experiment, measurement not available (univariate regression)
x_p	empirical quantile for probability p
Y	response variable
$\mathbf{y}$	observation vector of Y
y	observation of a response variable
y_L	particular limit of the response variable Y
y_+	observation of a response variable when input factor is at level "+"
y_-	observation of a response variable when input factor is at level "−"
$\hat{y}_*$	prediction of an experiment $\mathbf{x}_*$ (univariate regression)
$\boldsymbol{\mathcal{Y}}_t^0$	input matrix of an aging model
$\mathbf{Z}$	random variable modeling usage behavior
Z	random variable regarding the usage profiles considered
$\mathbf{z}$	usage behavior
z	usage profile
ANOVA	Analysis of Variance
ASC	Ammonia Slip Catalyst
ASQ	American Society for Quality

AVL	Anstalt für Verbrennungskraftmaschinen List
CAC	Charge Air Cooler
CF	Conformity Factor
CH$_4$	Methane
CI	Compression Ignition
CLT	Central Limit Theorem
CoP	Conformity of Production
DEF	Diesel Exhaust Fluid
DF	Deterioration Factor
DOC	Diesel Oxidation Catalyst
D-O-Q-R	Depth-Outlyingness-Quantile-Rank
DPF	Diesel Particulate Filter
DS-design	Definitive Screening Design
EAS	Exhaust Aftertreatment System
EC	European Commission
ECU	Engine Control Unit
EEC	European Economic Community
EGR	Exhaust Gas Recirculation
EOQ	European Organization for Quality
GPF	Gasoline Particulate Filter
H$_2$O	Water
HC	Hydrocarbon
HiL	Hardware in the Loop
IMSE	Integrated Mean Squared Error
iid	Independent and Identically Distributed
ind	Independently Distributed
ISC	In-Service Conformity
ISO	International Organization for Standardization
IQR	Interquartile Range
LCC	Life Cycle Costing
LCCs	Life Cycle Costs
LSL	Lower Specification Limit
LH-design	Latin Hypercube Design
MAD	median absolute deviation

MARS	Multivariate Adaptive Regression Splines
MSE	Mean Squared Error
MC	Monte Carlo (Simulations)
MIL	Malfunction Indicator Light
MiL	Model in the Loop
MLP	Multilayer Perceptron
MoBEO	Model Based Engine Optimization (Semi-Physical Simulation Tool)
N_2	Nitrogen Gas
NEDC	New European Driving Cycle
NMHC	Nonmethane Hydrocarbons
NH_3	Ammonia
NO	Nitric Oxide
NO_2	Nitrogen Dioxide
NO_x	Nitrogen Oxides ($NO + NO_2$)
NRTC	Nonroad Transient Cycle
OBD	On-Board Diagnostic
OEM	Original Equipment Manufacturer
PDCA	Plan-Do-Check-Act (Cycle)
PEMS	Portable Emission Measurement System
PGM	Platinum Group Metals
PNLTR	Japanese Post-New Long Term Regulations
PI	Positive Ignition
Pt	Platinum
QQ-Plot	Quantile Quantile Plot
RMSE	Root Mean Squared Error
RPD	Robust Parameter Design
SCR	Selective Catalytic Reduction
SDPF	SCR coated DPF
std	flag for standardized data
SPC	Statistical Process Control
TCO	Total Costs of Ownership
TSI	Total Sensitivity Index
TWC	Three Way Catalyst
UN/ECE	Economic Commission for Europe of the United Nations

USL	Upper Specification Limit
WHSC	World Harmonized Stationary Cycle
WHTC	World Harmonized Transient Cycle

Bibliography

ABOURIZK, S. M., HALPIN, D. W. and WILSON, J. R. (1994). "Fitting Beta Distribution Based on Sample Data". *Journal of Construction Engineering and Management*, Vol.120, pp. 288-305

ANDERSON, T. W. and DARLING, D. A. (1952). "Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes". *The Annals of Mathematical Statistics*, Vol.23, pp. 193-212

ANDERSON, T. W. (1962). "On the Distribution of the Two-Sample Cramer-von Mises Criterion". *The Annals of Mathematical Statistics*, Vol.33, pp. 1148-1159

ANDERSON, T. W. (1984). "*An Introduction to Multivariate Statistical Analysis*". 2nd Ed., New York, NY: John Wiley & Sons

BARBER, C. B., DOBKIN, D. P. and HUHDANPAA, H. T. (1996). "The Quickhull Algorithm for Convex Hulls". *ACM Transactions on Mathematical Software (TOMS)*, Vol.22, pp. 469-483

BISHOP, C. M. (1995). "*Neural Networks for Pattern Recognition*". 1st Ed., Oxford: Oxford University Press

BISHOP, C. M. (2007). "*Pattern Recognition and Machine Learning*". 1st Ed., New York, NY: Springer

BOX, G. E. P. and BEHNKEN, D.W. (1960). "Some New Three Level Designs for the Study of Quantitative Variables". *Technometrics*, Vol.2, pp. 455-476

BOX, G. E. P. and COX, D.R. (1964). "An Analysis of Transformations". *Journal of the Royal Statistical Society*, Vol.26, pp. 211-252

BOX, G. E. P. and WILSON, K.B. (1951). "On the Experimental Attainment of Optimum Conditions". *Journal of the Royal Statistical Society*, Vol.13, pp. 1-45

BREUSCH, T. S. and PAGAN, A. R. (1979). "A Simple Test for Heteroscedasticity and Random Coefficient Variation". *Econometrica: Journal of the Econometric Society*, pp. 1287-1294

BRUSSEE, W. (2012). "*Statistics for Six Sigma*". 2nd Ed., New York, NY: McGraw-Hill Professional

BYRD, R. H., LU, P., NOCEDAL, J. and ZHU, C. (1995). "A Limited Memory Algorithm for Bound Constrained Optimization". *SIAM J. Scientific Computing*, Vol.16, pp. 1190-1208

CASELLA, G. and BERGER, R. L. (2006). "*Statistical Inference*". 6th Ed., Pacific Grove, CA: Duxbury

CHEN, S., COWAN, C. F. N. and GRANT, P. M. (1991). "Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks". *IEEE Transactions on Neural Networks*, Vol.2, pp. 302-309

CHENG, S.-W. and WU, C.F.J. (2001). "Factor Screening and Response Surface Exploration (with Discussion)". *Statistica Sinica*, Vol.11, pp. 553-604

CONOVER, W. J. JOHNSON, M. E. and JOHNSON, M. M. (1981). "A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data". *Technometrics*, Vol.23, pp. 351-361

CROUX, C., FILZMOSER, P. and FRITZ, H. (2013). "Robust Sparse Principal Component Analysis". *Technometrics*, Vol.55, pp. 202-214

CURRIN, C., MITCHELL, T., MORRIS, M. and YLVISAKER, D. (1991). "Bayesian Prediction of Deterministic Functions: With Applications to the Design And Analysis of Computer Experiments". *Journal of the American Statistical Association*, Vol.86, pp. 953-963

DAGPUNAR, J. (1988). "*Principles of Random Variate Generation*". 1st Ed., Oxford, UK: Clarendon Oxford Science Publications

Delphi Automotive PLC (2012), "*Worldwide Emissions Standards Heavy Duty and Off-Highway Vehicles*". 5725 Delphi Drive, Troy Michigan 48098-2815 USA

DIESTEL, R. (2010). "*Graph Theory (Graduate Texts in Mathematics)*". 4th Ed., New York, NY: Springer

DOBLER, H., DOLL, W., GÜNTER, W., IGNATOWITZ, E. and Vetter, R. (2003). "*Fachkunde Metall*". 54th Ed., pp. 40-47. Haan-Gruiten, NW: Europa Lehrmittel

DONG, Z., HU, W. and XUE, D. (1994). "New Production Cost-Tolerance Models for Tolerance Synthesis". *Journal of Engineering for Industry*, Vol.116, pp. 199-206

DUDA, R. O. and HART, P. E. (2000). "*Pattern Classification (Electrical & Electronics Engr)*". 2nd Ed., New York, NY: John Wiley & Sons

DURBIN, J. and WATSON, G. S. (1971). "Testing for Serial Correlation in Least Squares Regression.III". *Biometrika*, Vol.58, pp. 1-19

EHRLENSPIEL,K. KIEWERT, A. LINDEMANN, U. (2014). "*Kostengünstig Entwickeln und Konstruieren: Kostenmanagement bei der Integrierten Produktentwicklung*". 7th Ed., Berlin, Heidelberg: Springer

ELLRAM, M. L. (1995). "Total Cost of Ownership: An Analysis Approach for Purchasing". *International Journal of Physical Distribution & Logistics Management*, Vol.25, pp. 4-23

FAHRMEIR, L., KÜNSTLER, R., PIGEOT, I. and TUTZ, G. (2010). "*Statistik*". 7th Ed., New York, NY: Springer

FANG, K. T. (1980). "The Uniform Design: Application of Number-Theoretic Methods in Experimental Design". *Acta Math. Appl. Sinica.*, Vol.3, pp. 363-372

FISCHER, B. (2011). "*Mechanical Tolerance Stackup and Analysis*". 2nd Ed., New York: CRC Press Inc

FISCHER, G. (2003). "*Lineare Algebra*". 14th Ed., Wiesbaden, HE: Vieweg

FISHMAN, G. S. (1996). "*Monte Carlo: Concepts, Algorithms and Applications*". 1st Ed., New York, NY: Springer

FRADEN, J. (2010). "*Handbook of Modern Sensors: Physics, Designs, and Applications*". 4th Ed., New York, NY: Springer

FRIEDMAN, J. H. (1991). "Multivariate Adaptive Regression Splines". *The Annals of Statistics*, Vol.19, pp. 1-141

GAUSS, C. F. (1809). "*Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium Auctore*". Sumtibus Frid. Perthes et IH Besser

GEISDÖRFER, K. (2009). "*Total Cost of Ownership (TCO) und Life Cycle Costing (LCC): Einsatz und Modelle: Ein Vergleich zwischen Deutschland und USA*". 1st Ed., Münster: LIT Verlag

GEORGE, L. M. PRICE, M. ROWLANDS, D and MAXEY J. (2004). "*The Lean Six Sigma Pocket Toolbook*". 1st Ed., New York, NY: McGraw-Hill

GÖTZE, U. and WEBER, T. (2008). "Total Cost of Ownership". *Zeitschrift für Planung & Unternehmenessteuerung*, Vol.19, pp. 249-257

GOLDFELD, S. M. and QUANDT, R. E. (1965). "Some Tests for Homoscedasticity". *Journal of the American Statistical Association*, Vol.60, pp. 539-547

GOOS, P. and JONES, B. (2011). "*Optimal Design of Experiments*". 1st Ed., The Atrium, Southern Gate, Chichester, West Sussex, UK: John Wiley & Sons

HAGAN, M. T. and MENHAJ, M. (1993). "Training Feed-Forward Networks with the Marquardt Algorithm". *IEEE Transactions on Neural Networks*, Vol.5, pp. 989-993

HAHN, G. J. and SHAPIRO, S. S. (1967). "*Statistical Models in Engineering*". 1st Ed., New York, NY: John Wiley & Sons

HECK, R. M. FARRAUTO, R. J. and GULATI S. T. (2009). "*Catalytic Air Pollution Control: Commercial Technology*". 3rd Ed., New Jersey: John Wiley & Sons

HEUSER, H. (2003). "*Lehrbuch der Analysis*". 15th Ed., Stuttgart, Leipzig, Wiesbaden: Teubner

HIEBL, J. REISENHOFER, S. AUER I., BÖHM, R. and SCHÖNER, W. (2010). "Multi-Methodical Realisation of Austrian Climate Maps for 1971 to 2000". *Advances in Science and Research*, Vol.6, pp. 19-26

HINKELMAN, K. and KEMPTHORNE O. (1994). "*Design and Analysis of Experiments*". 1st Ed., Chichester: John Wiley & Sons

HOCHMUTH, R. and MEERKAMM, H. and SCHWEIGER, W. (2007). "An Approach to a General View on Tolerances in Mechanical Engineering". *Journal of Quality Technology*, Vol.43, pp. 1-9

HOMMA, T. and SALTELLI, A. (1996). "Importance Measures in Global Sensitivity Analysis of Nonlinear Models". *Reliability Engineering and System Safety*, Vol.52, pp. 1-17

HYNDMAN, R. J. and FAN, Y. (1996). "Sample Quantiles in Statistical Packages". *American Statistician*, Vol.50, pp. 361-365

JOHNSON, N. L. and KOTZ S. (1976). "*Distributions in Statistics: Continuous Multivariate Distributions*". 1st Ed., New York: John Wiley & Sons

JOHNSON, M. E. (1987). "*Multivariate Statistical Simulation*". 1st Ed., New York, Chichester: John Wiley & Sons

JOHNSON, M. E., MOORE, L. M. and YLVISAKER, D. (1990). "Minimax and Maximin Distance Designs". *Journal of Statistical Planning and Inference*, Vol.26, pp. 131-148

JONES, B. and NACHTSHEIM, C.J. (2011). "A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects". *Journal of Quality Technology*, Vol.43, pp. 1-15

KARATZOGLOU, A. SMOLA, A. HORNIK and K. ZEILEIS, A. (2006). "kernlab: Kernel Methods Lab. R Package Version 0.8-1". CRAN.R-project.org

KERMAN, J. (2011). "A Closed-Form Approximation for the Median of the Beta Distribution". *arXiv:1111.0433v1, [math.ST]*, pp. 1-6

KIRSCHBAUM, E. "Entwicklung einer Statistischen Methodik zur Charakterisierung von Lastkollektiven". Master Thesis, Graz University of Technology

KOTZ, S. BALAKRISHNAN, N. and JOHNSON, N. L. (2000). "*Continuous Multivariate Distributions*". 1st Ed., New York: John Wiley & Sons

LEGENDRE, A. M. (1805). "Nouvelles Méthodes Pour La Détermination Des Orbites Des Comètes". *Courcier*, Paris, Appendix on least squares

LIN, S.-S., WANG, H.-P. and ZHANG, C. (1997). "Statistical Tolerance Analysis based on Beta Distributions". *Journal of Manufacturing Systems*, Vol.16/2, pp. 150-158

LIU, R. Y., PARELIUS, J. M., JESSE, M. and SINGH, K. (1999). "Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference (With Discussion)". *Annals of Statistics*, Vol.27, pp. 783-858

McCULLOCH, W. S. and PITTS, W. (1943). "A Logical Calculus of the Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics*, Vol.5, pp. 115-133

McKAY, D. J., CONOVER, W. J. and BECKMAN, R. J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code". *Technometrics*, Vol.21, pp. 239-245

MEE, R. W. (2009). "*Factorial Two-Level Experimentation*". 1st Ed., New York, NY: Springer

MERCER, J. (1909) "Functions of Positive and Negative Type and Their Connection with the Theory of Integral Equations". *Philosophical Transactions of the Royal Society A*, Vol.209, pp. 415-446

MICCHELLI, C. A. (1986). "Interpolation of Scattered Data: Distance Matrices and Conditionally Positive Definite Functions". *Constructive Approximations*, Vol.2, pp. 11-22

MISCHKE, C. R. (1980). "*Mathematical Model Building: An Introduction to Engineering*". 2nd Ed., Iowa, IA: Iowa State Pr.

MONTGOMERY, D. C. (2012). "*Introduction to Statistical Quality Control*". 7th Ed., New York, NY: John Wiley & Sons

MONTGOMERY, D. C. (2013). "*Design and Analysis of Experiments*". 8th Ed., New York, NY: John Wiley & Sons

MOODY, J and DARKEN, C. J. (1989). "Fast Learning in Networks of Locally-Tuned Processing Units". *Neural Computation*, Vol.1, pp. 281-294

MOSLER, K. (2013). "Depth Statistics". *Robustness and Complex Data Structures*, Berlin-Heidelberg: Springer

MOST, T. and WILL, J. (2008). "Metamodel of Optimal Prognosis-An Automatic Approach for Variable Reduction and Optimal Metamodel Selection". *Proc. Weimarer Optimierungs-und Stochastiktage*, Vol.5, pp. 1-22.

NGUYEN, D. and WIDROW, B. (1990). "Improving the Learning Speed of 2-Layer Neural Networks by Choosing Initial Values of the Adaptive Weights". *Proceedings of the International Joint Conference on Neural Networks*, Vol.3, pp. 21-26.

PIFFL, M. and STADLOBER, E. (2015). "Efficient Treatment of High Dimensional Computer Experiments". *Journal of Statistical Planning and Inference*, Vol.164, pp. 10-26

PINDYCK, R. S. and RUBINFELD, D. L. (2003). "*Mikroökonmie*". 5th Ed., München (Germany): Pearson Studium

POWELL, M. J. D. (1987). "*Algorithms of Approximation*". 2nd Ed., New York, NY: Clarendon Press

PRAUTZSCH, H. BÖHM, W. and PALUSZNY M. (2001). "*Bezier and B-Spline Techniques*". 1st Ed., Berlin: Springer

RAO, C. R. (1988). "Methodology Based on the L_1 Norm in Statistical Inference". *Sankhya Ser. A*, Vol.50, pp. 289-313

REIF, K. (2012). "*Sensoren im Kraftfahrzeug*". 2nd Ed., Wiesbaden, HE: Vieweg+Teubner Verlag

RINNE, H. (2009). "*The Weibull Distribution*". 1st Ed., Boca Raton: Chapman and Hall CRC

RINNE, H. and MITTAG, H. J. (1995). "*Statistische Methoden der Qualitätssicherung*". 3rd Ed., München: Carl Hanser Verlag

ROHATGI (1976). "*An Introduction to Probability Theory and Mathematical Statistics*". 1st Ed., New York, NY: John Wiley and Sons

ROOS, D. MOST, T. UNGER, J. F. WILL, J (2007). "Advanced Surrogate Models within the Robustness Evaluation". *In Proceedings of the Weimarer Optimierungs- und Stochastiktage 4.0*, pp. 1-43

ROSENBLATT, F. (1962). "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms". Washington,*Spartan Books*

RUMELHART, D. E., HINTON, G. E. and WILLIAMS, R. J. (1986). "Learning Internal Representations by Error Propagation". *D. E. Rumelhart, McClelland, J. L. and the PDP Research Group (Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol.1, pp 318-362

SACHS, L., HEDDERICH, J. (2006). "*Angewandte Statistik*". 12th Ed., New York, NY: Springer

SACKS, J., SCHILLER, S. B. and WELCH, W. (1989). "Designs for Computer Experiments". *Technometrics*, Vol.31, pp 41-47

SCHOENBERG, I. J. (1973). "*Cardinal Spline Interpolation*". 12th Ed., Philadelphia, PA: Society for Industrial and Applied Mathematics

SCHOLZ, F. (1995). "Tolerance Stack Analysis". *Tolerance Stack Analysis Methods. Research and Technology Boeing Information & Support Services, Boeing Seattle*, pp. 1-44

SCHÜßLER, M., ALLMER, I., HOLLAUF, B., KORDON, M., KOZLIK, C., SEEWALD, G. and HÜLSER, H. (2008). "Computational Emission Optimization on Powertrain Level". *Aachener Kolloquium Fahrzeug- und Motorentechnik 2008*, pp. 1-26

SERFLING, R. (2010). "Equivariance and Invariance Properties of Multivariate Quantile

and Related Functions and the Role of Standardization". *Journal of Nonparametric Statistics*, Vol.22, pp. 915-936

SHAPIRO, S. S. and WILK, M. B. (1965). "An Analysis of Variance Test for Normality (for Complete Samples)". *Biometrika (1965)*, pp. 591-611

SHEWART, W. A. (1939). "Statistical Methods from the Viewpoint of Quality Control". Courier Corporation, 1939

SHEWRY, M. C. and WYNN, H. P. (1987). "Maximum Entropy Sampling". *Journal of Applied Statistics*, Vol.14, pp. 165-170

SIEBERTZ, K., VAN BEBBER, D. and HOCHKIRCHEN, T. (2010). "*Statistische Versuchsplanung*". 1st Ed., Berlin-Heidelberg: Springer

STEIN, M. (1987). "Large Sample Properties of Simulations Using Latin Hypercube Sampling". *Technometrics*, Vol.29, pp. 143-151

TAGUCHI, G. (1987). "*System of Experimental Design: Engineering Methods to Optimize Quality and Minimize Cost*". 1st Ed., White Plains, NY: UNIPUB

TAGUCHI, G. and WU, Y. (1980). "*Introduction to Off-Line Quality Control*". 1st Ed., Nagoya, Japan: Central Japan Quality Control Association

TAGUCHI, G. JUGULUM, R. and TAGUCHI, S. (2004). "*Computer-Based Robust Engineering*". 1st Ed., Milwaukee (USA): ASQ Quality Press Books

TIMISCHL, W. (1995). "*Qualitätssicherung*". 1st Ed., München, Wien: Hanser

TOUTENBURG, H. and KNÖFEL, P. (2008). "*Six Sigma*". 2nd Ed., New York, NY: Springer

TSAI, P.-W., GILMOUR, S. G. and MEAD, R. (2000). "Projective Three-Level Main Effects Designs Robust to Model Uncertainty". *Biometrika*, Vol.87, pp. 467-475

VAN LINT, J. H., SEIDEL, J. J. (1966). "Equilateral Point Sets in Elliptic Geometry". *Mathematical Sciences*, Vol.69, pp. 335-348

VENABLES, W. N. and RIPLEY, B. D. (2002). "*Modern Applied Statistics with S*". 4th Ed., New York, NY: Springer

WAUGH, J. G. (1961). "Evaluation of Integral of Elliptic Gaussian Distribution over a Centered Ellipse". *Naval Ordnance Test Station China Lake CA*, No. TP-2766, pp. 1-18

WEIBULL, W. (1939a). "A Statistical Theory of Strength of Materials". *Ingeniörs Vetenskaps Akademiens Handligar Report*, Vol.151, Stockholm

WEIBULL, W. (1939b). "The Phenomenon of Rupture in Solids". *Ingeniörs Vetenskaps Akademiens Handligar Report*, Vol.153, Stockholm

WIDROW, B. and HOFF, M. E. (1960). "Adaptive Switching Circuits". *In IRE WESCON Convention Record*, Vol.4, pp. 96-104

WU, J. C. F. and HAMADA, M. S. (2009). "*Experiments: Planing, Analysis, and Optimization*". 2nd Ed., New York, NY: John Wiley & Sons

XIAO, L., LIN, D. K. and BAI, F. (2012). "Constructing Definitive Screening Designs Using Conference Matrices". *Journal of Quality Technology*, Vol.44, pp. 1-7

XU, H. (2005). "A Catalogue of Three-Level Fractional Factorial Designs". *Metrika*, Vol.62, pp. 259-281

YULE, U. (1897). "On the Theory of Correlation". *Journal of the Royal Statistical Society*, Vol.60, pp. 812-854

ZEHBOLD, C. (1996b). "Lebenszykluskostenrechnung". *Dissertation, Universität Erlangen-Nürnberg*

ZUO, Y. and SERFLING, R. (2000). "General Notions of Statistical Depth Function". *Annals of Statistics*, Vol.28, pp. 461-482