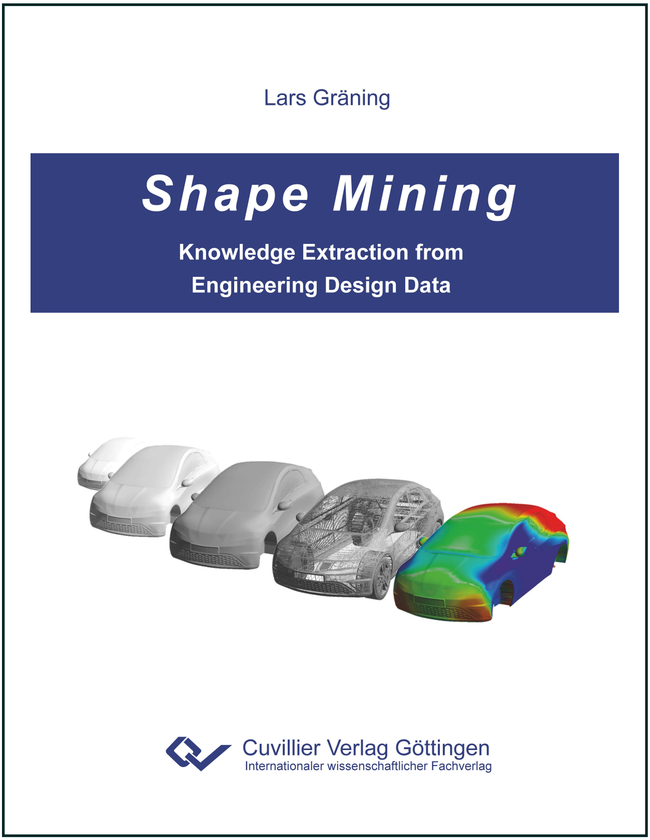
Lars Gräning (Autor)

Shape Mining

Knowledge Extraction from Engineering Design Data



Lars Gräning

Shape Mining

Knowledge Extraction from
Engineering Design Data

Cuvillier Verlag Göttingen
Internationaler wissenschaftlicher Fachverlag

https://cuvillier.de/de/shop/publications/6795

Copyright:
Cuvillier Verlag, Inhaberin Annette Jentzsch-Cuvillier, Nonnenstieg 8, 37075 Göttingen, Germany
Telefon: +49 (0)551 54724-0, E-Mail: info@cuvillier.de, Website: https://cuvillier.de

1

Introduction

1.1 Motivation

Computer Aided Engineering (CAE) technologies have become powerful tools especially but not limited to the design, evaluation and verification of technical systems. The modeling of physical processes and configurations on computer systems constitutes a major basis for an efficient product development process and for the solution of a huge variety of engineering problems. Technologies for Computer Aided Design (CAD), Computational Fluid Dynamics (CFD) simulation, Finite Element Analysis (FEA), computational optimization and response surface methodologies are well established in the domain of engineering design, and are gradually replacing expensive physical modeling processes. An increase of data resulting from this boost in the usage of computational engineering tools together with high pressure toward faster and efficient product development cycles makes a consistent and thorough information and knowledge handling imperative.

More recently, technologies from computational intelligence and data mining have been adopted to exploit experimental design data and computational resources for the support of engineers in the decision making process. However, the multidisciplinary characteristics of complex design processes and the huge variability in computational design representations hinders the analysis of design data beyond individual design configurations and processes. Especially the variation in the computational representations being used, makes an efficient knowledge exchange between various design processes difficult.

The fusion of information from different computational models becomes indispensable for implementing a holistic data mining process and goes beyond the storage and administration of raw design data. It requires the definition of a unified and typically high dimensional system design representation. Such a representation makes designs exchangeable between design processes and engineers and builds the basis for a holistic analysis of the design data.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Deploying sophisticated technologies from statistics, information theory, data mining, and computational intelligence upon a unified object representation is assumed to be the key paradigm for improving the efficiency of the design process, and for reducing the time to market for products in the future. With the integration of data mining technologies on a unified basis, decisions and knowledge can be made indelible throughout the existence of the developed design and beyond potential reorganizations of design processes, engineering teams and competences.

However, the application of data mining or related technologies to a high dimensional unified representation goes beyond the usage of modern data modeling techniques, it requires the consideration of all aspects of the data mining process including pre-processing, feature extraction, data modeling, visualization and utilization of the considered design data.

The concepts for *shape mining*, introduced in this thesis, instantiate unstructured surface meshes as a unified representation of the shape of three dimensional objects related to the domain of aerodynamic and structural design. The described concepts focus on the investigation of technical systems designed for optimal fluid dynamics performance, motivated from the request of engineers in the domain of car, airplane and boat design. Techniques and prerequisites for integrating technologies regarding the analysis of shape data and the support for decision making within the engineering design process are studied. The suggested methodologies consider the representation of designs, the extraction of valuable information from existing design and performance data and the utilization of the acquired information.

1.2 Overview

An overview of the *shape mining* concept is depicted in Fig. 1.1. The diagram summarizes the studied topics of this thesis, which is organized as follows.

Chapter 2 provides an introduction to the engineering design synthesis process. The interplay between engineers, design synthesis and analysis are discussed on a conceptual basis. Thereafter, a compact review of the state of the art in computer aided engineering and design analysis is given. As an example, aspects of the synthesis process are discussed based on the generation of passenger car design data. The resulting designs and performance numbers are the basis for experimental studies throughout the thesis.

Chapter 3 introduces unstructured surface meshes as a unified object representation in the domain of aerodynamic and structural design. The definition of a unified object representation is a pre-requisite for the implementation of a holistic data modeling process. Methods for evaluating

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

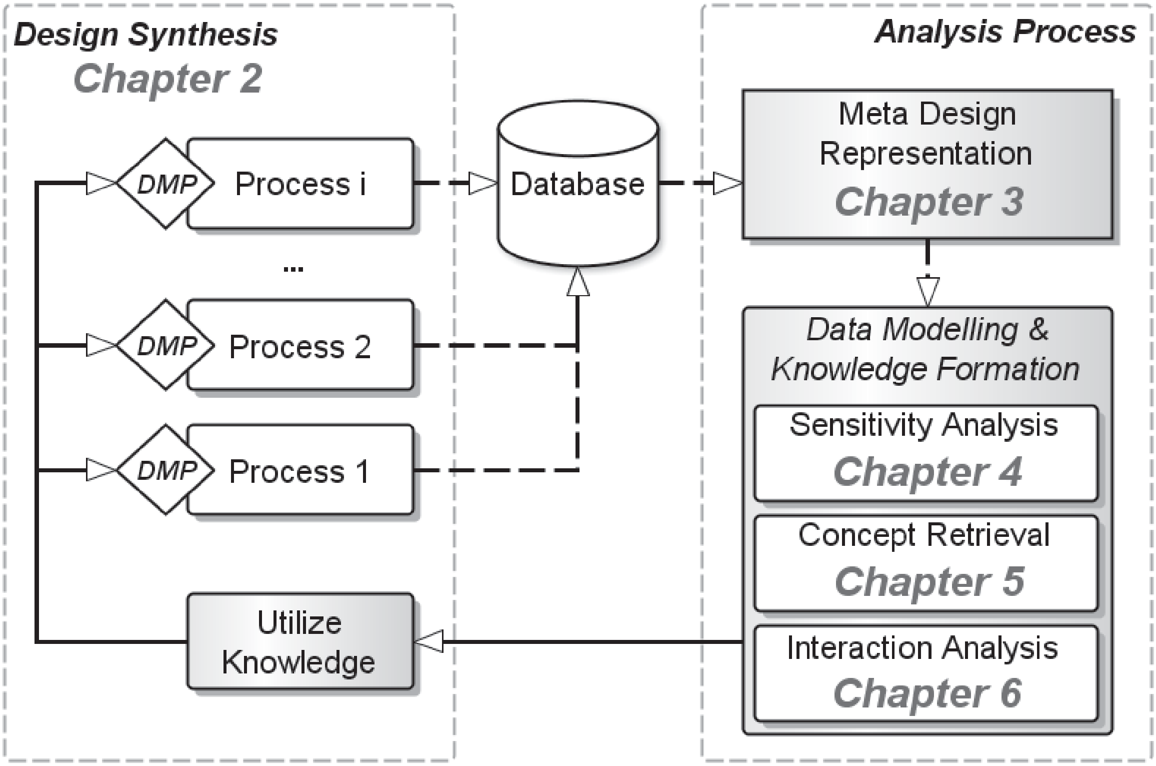


Figure 1.1: *Low level view on the shape mining process. Summary of the main components of the thesis.*

the differences between local surface characteristics are described. Based on the surface mesh representations of the passenger car designs, a statistical analysis of the surface feature variations provides means, e.g., to compare individual designs or to evaluate the course of design processes.

Chapter 4 studies methods for extracting knowledge about the interrelation between design feature variations and changes in the design performance based on the unified shape representation. Standard methods for sensitivity analysis are reviewed and a robust variant of the mutual information is compared to its information theoretic definition. Applied to the passenger car design data, the sensitivity analysis provides tools to filter potentially irrelevant design features from the design data. K-nearest neighbor algorithms are studied, allowing to control the locality of the sensitivity estimates. The application of more sophisticated methods of knowledge formation necessitates to resolve the typical drawback of the universal design representation, i.e., its high dimensionality. Exploiting the geodesic distance together with the sensitivity information, a new algorithm is proposed to derive larger sensitive design areas, and with that derive a reduced object shape representation.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Chapter 5 investigates the retrieval of abstract design concepts. The reduced feature sets from the sensitivity analysis facilitate the application of enhanced modeling techniques from data mining and computational intelligence to the design data. In this chapter, the procedure for the retrieval, representation and evaluation of abstract design concepts is generalized and can be carried out independently of the used modeling technique. A new measure of concept relevance is defined, which evaluates extracted design concepts based on the estimation of their utility. The new measure allows the ranking of concepts according to the formulation of the engineers' objectives. Self-organizing maps and decision tree models are studied and applied to the passenger car design data with the purpose of retrieving relevant design concepts described by human-readable design rules.

In **Chapter 6** an information theoretic approach for the identification of design interactions is investigated. The measure of interaction information, an extension of the mutual information to multiple variables, provides a computational means to quantify the joint influence of distant design areas on the design performance. The resulting interaction graph, a graphical representation of the information measures, provides a fast and easy way to visualize rather complex statistical dependencies. Results from the interaction analysis provide valuable information, which can be utilized to decompose complex design tasks into simpler, functional independent subcomponents.

The major results and contributions of this thesis are summarized in **Chapter 7**. Ideas and possible directions for future research are proposed in the outlook of the thesis.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

2

Design Synthesis

The engineering design synthesis is a versatile process with a close cooperation between engineers, experimental facilities and computer systems, which furthermore requires an efficient interplay between design and analysis activities, conceptually discussed in this chapter. In the recent years, many computational tools and methodologies have been developed to support engineers in the design, optimization and analysis of technical systems. A compact overview on the state of the art methods in computer aided engineering design are given in this chapter.

The generation of computer readable design data during the design synthesis, e.g., using computer aided engineering tools, is the groundwork for a holistic data driven analysis and knowledge extraction framework. Given the exterior design of a passenger car as an example, relevant properties of the design process are explained. Throughout the thesis, the generated passenger car design data defines a solid base for studying data mining concepts for knowledge extraction within the shape mining framework.

2.1 Engineering Design

The engineering design process can best be described as a goal oriented iterative decision making process [Pahl et al., 2007; Evbuomwan et al., 1996]. In each iteration, engineers decide about individual or a sequence of design variations that lead to a final design configuration, fulfilling pre-defined constraints and design goals. This requires that initial decisions about the design representation, the evaluation technique and constraints are made, which are continuously reviewed during the progress of the design process. The efficiency and success of any design process depends on the decisions taken and thus on the experiences and knowledge of the engineers. In order to augment and distribute any supplementary knowledge, a distinct analysis process is indispensable after the actual modeling, production and evalua-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

tion of design instances. As depicted in Fig. 2.1, each design process can be partitioned into distinct phases: the actual *synthesis process*, referring to the process of generating new design variations, the *analysis process*, relating to the formation of new domain knowledge from the results, and finally the *decision making process* that derives new formulations about the design strategy, representation, constraints and objectives.

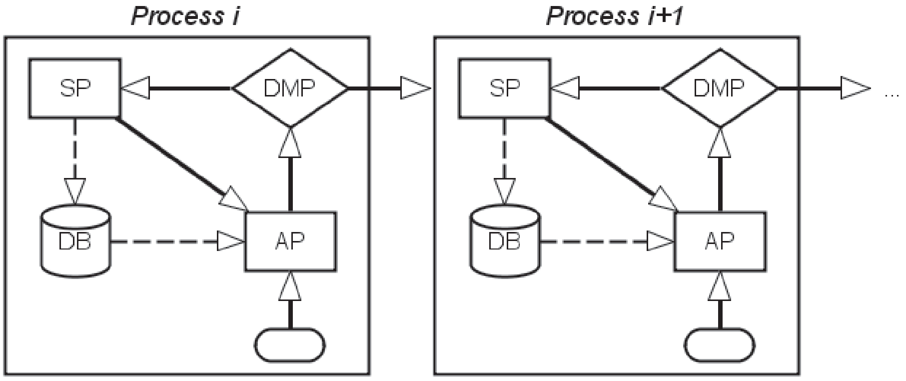


Figure 2.1: *Flow chart of a serial design process configuration. Process i defines the baseline of the subsequent process $i+1$, where design information is passed from process i to its successor. Each process includes the actual synthesis process (SP), the analysis processes (AP), the decision making process (DMP) and a database (DB) for data storage.*

Large scale systems often require a subdivision of the entire design configuration and the design process in order to handle their complexity. Thus, the outcome of each design process defines just an intermediate step within the overall synthesis process. Such interrelations between processes can be classified into serial or parallel process configurations, illustrated in Fig. 2.1 and Fig. 2.2, respectively.

In a serial configuration, preceding design processes define the starting configuration or constraints for a subsequent design process. Process information is transferred at the interface between successive processes. Typically, the exchange of information is limited to information related to the final design configuration. The exploitation of intra-process information from process i by process $i+1$ is not modeled in such a pattern and completely relies on the engineers involved. Thus, design data and the results from the analysis and decision making are properties of each individual process, where in each process often distinct representations of the design or parts of it are used. Hence, variations are applied and decisions are made based on individual representations, ignoring the holistic view on the design. The analysis

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

process, where decisions are made, refers to the representations of the design in the process and does not integrate information, e.g., from preceding processes.

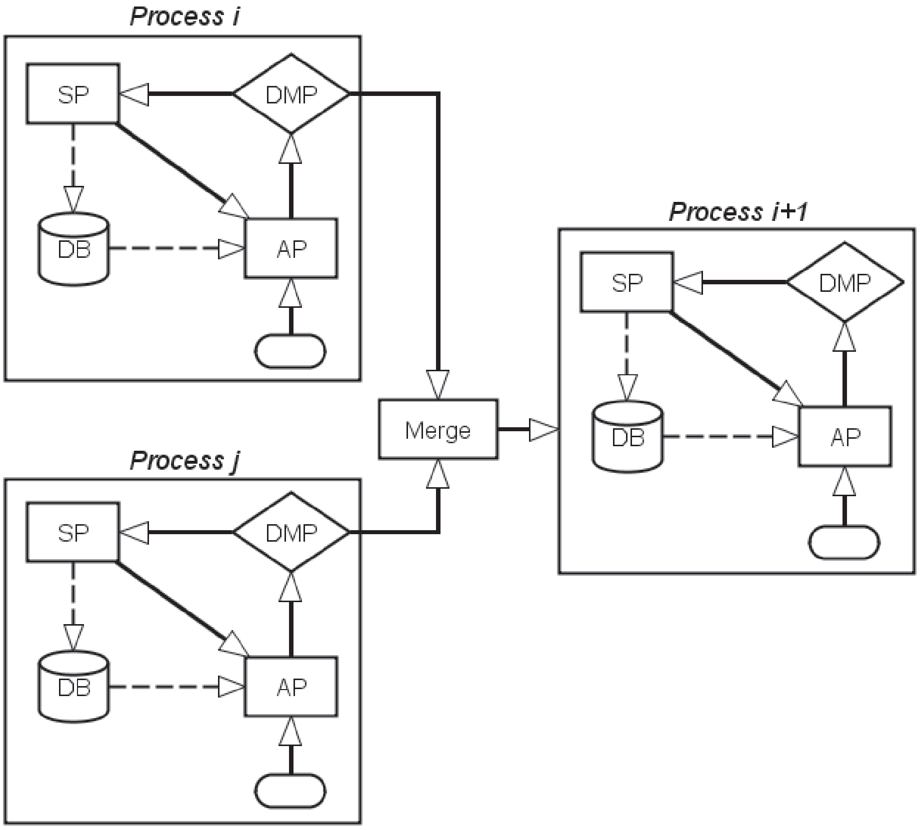


Figure 2.2: *Flow chart of a parallel process configuration. Results of parallel processes* i *and* j *need to be merged before starting a subsequent process* $i+1$*. Each process includes the actual synthesis process (SP), the analysis processes (AP), the decision making process (DMP) and a database (DB) for data storage.*

In a parallel design configuration a design process i is split into parallel sub-processes i and j, as depicted in Fig. 2.2. The parallelization of individual design processes can increase the efficiency of the overall process. However, results of parallel processes need to be merged before starting any subsequent process. Thereafter, design information is exchanged after finalizing all parallel processes. A design configuration which is superior with respect to the constraints and targets of one design process i might be inferior in the light of a second concurrent process j. In an extreme scenario, different parallel design processes might even act against each other, especially when each

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

process targets the design of a distinct part of the holistic design. Merging the results of parallel processes can sometimes lead to surprising observations with respect to overall design properties. Additional expensive subsequent design steps have to be carried out to balance possible debasements.

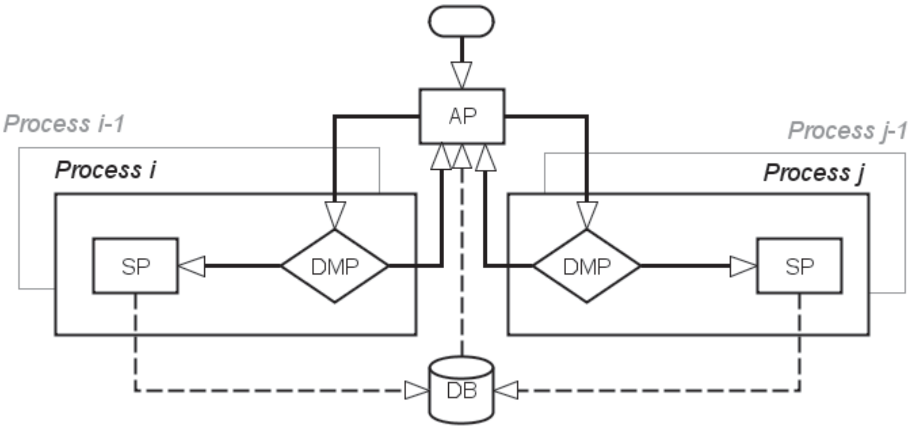


Figure 2.3: *Flow chart of the integrative design processes configuration. The synthesis process (SP) and the decision making process (DMP) are properties of individual processes, whereas the analysis process (AP) integrates information from all concurrent and serial processes, requiring the instantiation of a shared database (DB).*

Both, serial and parallel configurations are found in nearly any real world design process, even so they are not implemented in its pure form. Processes might partially overlap. Furthermore, serial and parallel process patterns are used in combination. However, the problems described above still hold. Analysis and decision making processes are mostly a property of individual design processes.

For an efficient co-evolution of design processes, design data and related information needs to be accumulated beyond individual processes throughout the course of the entire design synthesis, by means of implementing an integrative process configuration. The concept of the integrative process configuration is depicted in Fig. 2.3. The implementation of a shared database together with a unification of the design data, i.e., the definition of a unified design representation and the unified evaluation of the design performance, facilitates the implementation of a holistic analysis process. In such a configuration, individual design decisions can be made in consideration of the holistic design. Such global and holistic view on the design data allows an integrated exchange of domain knowledge at any stage of the design pro-

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

cess. Thus, a late merging of process results could be omitted, increasing the efficiency of the overall process.

To exploit the potential of the integrative process configuration, methods from the Computer Aided Engineering (CAE) and Data Mining domain need to be utilized to aid the analysis and design synthesis process. Furthermore, new attempts for the holistic analysis of the design data need to be established.

2.2 Computer Aided Design Synthesis

Engineers involved in the synthesis of new designs are supported by a huge assortment of Computer Aided Engineering (CAE) tools, which aid, e.g., the modeling, creation, evaluation and optimization of technical systems. With the introduction of Computer Aided Design (CAD) systems, engineers moved from their writing desk to computerized workstations [Keane and Nair, 2005]. Initially designed for drafting, nowadays, modern CAD systems capture information about geometric and material properties, as well as process and manufacturing information for individual designs. From the specification of the design shapes, high fidelity simulation tools for Computational Fluid Dynamics (CFD) and Finite Element Analysis (FEA) facilitate the evaluation of the functional characteristics of each design on the computer, step by step replacing expensive physical experiments. With the support of advanced visualization technologies, domain experts can profit from the analysis of individual simulation results by accumulating information for consolidated decision making. Based on the design constraints, the objectives of the design process and the result of individual simulations, engineers decide about successive design configurations.

While the decision making based on individual design and evaluation results is a demanding process, reflecting a trial and success attempt, methods for the Design of Experiments (DOE) [Forrester et al., 2008; Morris and Mitchell, 1995] allow a more systematic approach to explore feasible design configurations. On the basis of the DOE results, response surface methodologies [Myers et al., 2009] are applicable to speed up the experimental design process by identifying potentially outperforming design configurations. The combination of computational optimization technologies with high fidelity simulations for CFD simulations or FEA can be seen as an additional big step towards the automation of design processes. Given a set of pre-defined constraints and objective settings, and typically given a fixed object representation, computational optimization technologies implement heuristics to imitate the iterative decision making process on computers, see [Keane and

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.

Nair, 2005] for an introduction. Nature inspired, derivative free search strategies like genetic algorithms [Holland, 1992], evolution strategies [Rechenberg, 1971; Beyer and Schwefel, 2002], simulated annealing [Kirkpatrick et al., 1983] or particle swarm optimization [Kennedy and Eberhart, 1995] became prominent in the domain of engineering design in the recent years. An extensive overview concerning such optimization technologies is provided by [Montaño et al., 2012] or [Saridakis and Dentsoras, 2008].

With the selection of the computational optimization algorithm, and with the specification of the representation, optimization objectives and constraints, the engineers define the strategy for each individual design process. Computer aided optimization methods, e.g., the CMAES (Evolution Strategy with Co-variance Matrix Adaptation) [Hansen and Ostermeier, 2001] or model guided evolution strategies [Graening et al., 2010; Reehuis et al., 2011] even adapt their search strategy online depending on the current state of the search process. As shown in [Olhofer et al., 2001], an online adaptation of the design representation during the search does relax the initial search constraints and can lead to new outperforming design solutions.

In engineering design optimization the evaluation of the objective functions often requires to conduct computational expensive high fidelity simulations, limiting optimization and search strategies to unfold their theoretic potential. Surrogate assisted search strategies, as the result of the progress made in the domain of artificial intelligence, allow a partial replacement of the expensive fitness evaluations by fast approximation models during the progress of the optimization, e.g., see [Jin, 2005; Forrester et al., 2008; Jin, 2011] for an overview.

Each search and optimization process aided by computer systems produces a multitude of computer readable design data. Beyond the optimized design configuration, each design and its related performance carries important information about the design domain. The extraction of valuable information from all the generated design data requires a systematic attempt for data analysis.

2.3 Data Analysis

The analysis of design data generated is an indispensable step in any design process. A precise data analysis may produce new domain knowledge, which can support engineers in decision making. Methods from data mining provide a systematic formalism and the necessary means for the analysis of stocks of data.

Dieses Werk ist copyrightgeschützt und darf in keiner Form vervielfältigt werden noch an Dritte weitergegeben werden.
Es gilt nur für den persönlichen Gebrauch.