

---

# Sparse Data Enrichment by Context Oriented Model Reduction Techniques in Manufacturing Industry with an Example Laser Drilling Process

---

You Wang<sup>1</sup>, Hasan Tercan<sup>2</sup>, Torsten Hermanns<sup>3</sup>, Thomas Thiele<sup>2</sup>,  
Tobias Meisen<sup>2</sup>, Sabina Jeschke<sup>2</sup> and Wolfgang Schulz<sup>1,3</sup>

<sup>1</sup>*Nonlinear Dynamics of Laser Manufacturing Processes Instruction and Research Department (NLD) of RWTH Aachen University, Steinbachstraße 15, 52074, Aachen, Germany*

<sup>2</sup>*Institute of Information Management in Mechanical Engineering (IMA) of RWTH Aachen University, Dennwartstraße 27, 52068, Aachen, Germany*

<sup>3</sup>*Fraunhofer Institute for Laser Technology, Steinbachstraße, 52074, Aachen, Germany*  
*E-mail: you.wang@ilt.fraunhofer.de*

Received 05 April 2018;

Accepted 23 July 2018

## Abstract

Nowadays, the internet of things and industry 4.0 from Germany are all focused on the application of data analytics and Artificial Intelligence to build the succeeding generation of manufacturing industry. In manufacturing planning and iterative designing process, the data-driven issues exist in the context of the purpose for approaching the optimal design and generating an explicit knowledge. The multi-physical phenomena, the time consuming comprehensive numerical simulation, and a limited number of experiments lead to the so-called sparse data problems or “curse of dimensionality”. In this work, an advanced technique using reduced models to enrich sparse data

*Journal of ICT, Vol. 6\_3, 203–216. River Publishers*

doi: 10.13052/jicts2245-800X.632

*This is an Open Access publication. © 2018 the Author(s). All rights reserved.*

is proposed and discussed. The validated reduced models, which are created by several model reduction techniques, are able to generate dense data within an acceptable time. Afterwards, machine learning and data analytics techniques are applied to extract unknown but useful knowledge from the dense data in the Virtual Production Intelligence (VPI) platform. The demonstrated example is a typical case from laser drilling process.

**Keywords:** sparse data, industry data, model reduction, machine learning, virtual production intelligence.

## 1 Introduction

Presently, manufacturing companies in the high-wage countries have to overwhelm the challenges of rapid responding to variant market demands, individual customer necessities and increasing labour costs. The rising complexity in the production processes motivates the generation of more reliable, efficient and flexible production planning and scheduling steps. The process parameter identification, knowledge extraction, iterative and communicative process design and multidisciplinary optimization are vital fields for production planning and decision making. Since a large amount of data are generated from machines, sensors, orders etc. in manufacturing industry, the developed data-driven methods and methodology can be applied in manufacturing decision making processes [1]. However, the data density is not enough for extracting knowledge because the evident parameter dimensionality is not enough and “the number of experiments” is limited by time or costs.

In this work, a methodology to enrich sparse data by fast and frugal reduced models is introduced. Several typical model reduction methods such as mathematical methods, numerical methods and data-driven methods for generating reduced models are reviewed. After obtaining sufficient and dense data, machine learning methods such as clustering and classification are applied to conduct the data analysis and knowledge extraction. The example case is a laser drilling process. The detailed enrichment of the data and the data-driven decision making process are demonstrated in this example.

## 2 Sparse Data Problems in Manufacturing Industry

The processes in manufacturing industry are characterized by multi-parameter models with high resolution. Moreover, the solvability of the process is considerably to be restricted by the complexity of physics. From a data

analytics perspective, two main barriers are addressed in this paper, which slow down the process to extract knowledge from manufacturing processes.

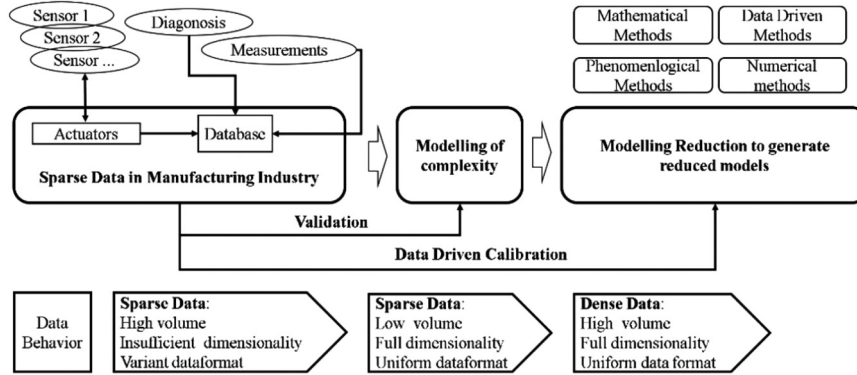
The first reason is that the required number of sampling points is enormous. When the dimensionality of parameters increases, the volume of sampling space increases exponentially. That means a large amount of parameters combination is needed and the existing available data becomes sparse. Especially, in manufacturing industry, the data from different process domains becomes heterogeneous and sparse within the high dimensional parameter space.

The second reason is that the sampling process can be time consuming. The sampling process is a process of selecting or generating the dataset which can be used in knowledge extraction. The sampling points can be collected from the real experiments or the computer aided calculations. The number of real experiments is not only limited by the time restriction but also limited by the boundary of the performance of machines or sensors. The computer aided calculation, also known as numerical simulation, is a powerful tool to generate sampling datasets. However, the complex numerical simulation could be time-consuming because of the high resolution and complex process physics.

### **3 Enriching Sparse Data by Reduced Models**

Generally speaking, the reduced models can enrich the sparse data from two aspects. On the one hand, the reduced models decrease the required data volume dramatically. The “sparse data” become “dense data” because of the compendious dimensionality. On the other hand, the reduced models are also characterized by their convenient solvability, millions of datasets can be generated from reduced models in acceptable duration.

The workflow to enrich sparse data into dense data are shown in Figure 1. Firstly, the data is extracted from divergent sources in real manufacturing processes and experimental measurements. Different types of sensors, diagnosis techniques and design of experiments are intensively operated to accumulate the original sparse data. At this stage, the data are featured by its high volume, insufficient dimensionality, heterogeneous distribution and irregular data format. Thereafter, the massive mathematical and physical modelling work for the complex manufacturing process is performed and validated by the sparse data in the first stage. From the well-built complicated models, the full dimensional parameters will be involved and the data generated from multifarious models is equipped with standard data format. Since the sampling process by complex models is time consuming, model reduction



**Figure 1** Workflow to enrich sparse data to dense data.

techniques are applied to generate the fast and frugal reduced models and avoid the unnecessary complexity. The reduced models are derived to avoid any unnecessary complexity and to reduce the computation time of large-scale dynamical systems by inducing approximations of much lower dimensions which can produce nearly the same input-output response characteristics. Meanwhile, to ensure the accuracy and the applicability on the specific context, the reduced models are also calibrated and validated by the measured sparse data. These reduced models can simulate the complex system by preventing redundant calculation, so the sparse data can be enriched dense enough for data-driven decision making extraction within a short phase.

The model reduction procedure adopts a top-to-down approach and starts from the original partial differential equations and derives approximated analytical solutions or a set of ordinary equations using many mathematical approaches, physical and phenomenological approaches, numerical approaches and data driven model reduction methods. Especially, there are several model reduction techniques which are convenient enough to use and worth reviewing.

### 3.1 Perturbation Analysis

The objective of perturbation theory is to determine the behaviour of the solution when one variable tends to be very small, which can lead to the split of two part of solutions for complex system. One part is the temporal solution and another part is the long term asymptotic solution. This separation of system solutions result in the reduction for the models. The typical application of this perturbation theory lies in the fields involving differential equations as well as

a series of engineering problems [2, 3]. Vossen et al. used the perturbation and asymptotic analysis to describe the dynamical behaviours of the free boundaries of the melt during the laser cutting process considering the spatially distributed laser radiation. A reduced model which can generate results at real time scale was derived by perturbation analysis for the purpose of predicting the product roughness [4].

### **3.2 Inertia and Central Manifold Analysis**

The inertial manifolds are connected with the long term behavior of the solutions of dissipative dynamical systems. The reduced phase space of ordinary differential equations and partial differential equations in the long time limit is named as central or inertial manifold. Schulz et al. derived a reduced model by applying inertial manifold method. This reduced model can calculate the thermal behavior in laser manufacturing processes very fast [5, 6].

### **3.3 Buckingham Pi Theory**

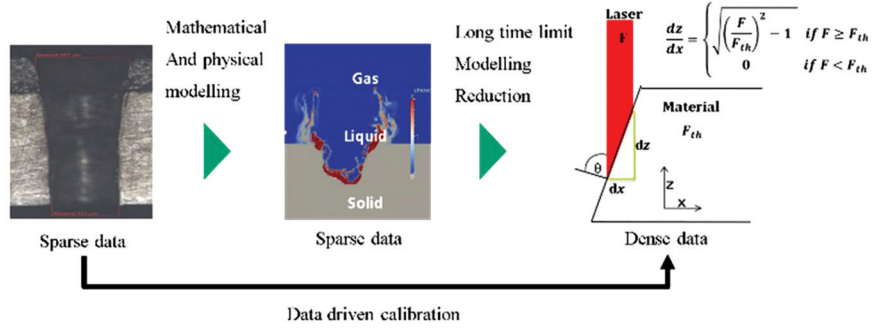
The Buckingham Pi theory is used to find the dimensionless groups from relevant input and output parameters. The dimensionality of the original parameters is sharply decreased and simplified by applying the dimensionless groups [7].

### **3.4 Proper Orthogonal Decomposition**

Proper Orthogonal Decomposition (POD) is a numeric method by searching for a low-dimensional approximate representation of the large scale dynamical systems, such as signal analysis, turbulent fluid flow and large dataset like image processing [9, 10]. POD generates a set of orthonormal basis of dimensions, which minimizes the error from approximating the snapshots. It can generally give a good approximation with substantially lower dimensionality [11].

## **4 Example for Laser Drilling Manufacturing**

As an example to illustrate the data enrichment by reduced models and data visualization, an advanced reduced model for sheet metal drilling has been developed by Nonlinear Dynamics of Laser Manufacturing Processes Instruction and Research Department (NLD) in RWTH Aachen University [12]. Using this model, the final shape of the drilling holes can be calculated



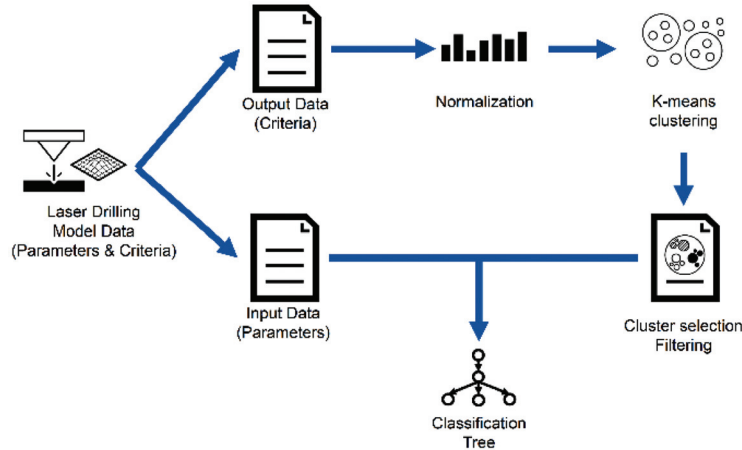
**Figure 2** The procedures to generate asymptotic laser drilling reduced model.

and described. Inside the formula (see Figure 2), the term  $F$  is the local laser fluency,  $z$  and  $x$  represent the position along  $z$  and  $x$  axis respectively, the term  $F_{th}$  is based on the heuristic concept of an ablation threshold and material dependency. The only one unknown parameter has to be calibrated and determined with experimental sparse data. Afterwards, this reduced model can be used to calculate the final shape of the drilling hole by laser sheet metal drilling. Not only the final shape of drilling hole but also the feasibility for each parameter can be indicated accurately by this reduced model.

As a consequence, the whole asymptotic shape of the drilling hole is calculated and is illustrated. Finally, classification of sheet metal drilling can be performed by identification of the parameter region where the drilling hole achieves its asymptotic shape.

The asymptotic drill reduced model is an ordinary differential equation which can be solved within 1 second for each single run. This fast reduced model makes it possible to collect dense data in an acceptable period. By using the asymptotic drill reduced model, 10,000 sampling points can be generated within five seconds without parallel calculations. However, it takes 30 minutes to produce one sample if the complicated numerical simulation is adopted.

After the sparse data is enriched into dense data by asymptotic reduced model, the machine learning techniques are applied to conduct data analytics. Thereby the data analytics process including appropriate data visualization methods are implemented within a Virtual Production Intelligence (VPI) platform [13]. The process is described in detail in [14]. It implemented a hybrid data analytics approach with clustering and classification tree to identify parameters of the manufacturing process that result in desired outputs. The approach is shown in Figure 3.



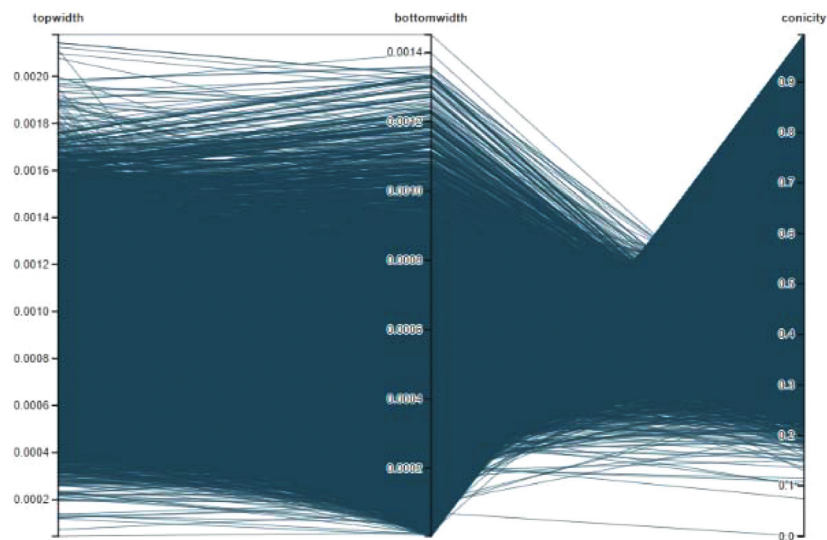
**Figure 3** Data analytics process to analyse dense data [14].

First of all, clustering is used to divide the output data of the reduced model into different regions (clusters). Thus, the user is able to select a cluster that represents desired process results. After that, the classification trees are used to identify regions of the parameters space (see also Table 1) which lead to these process result.

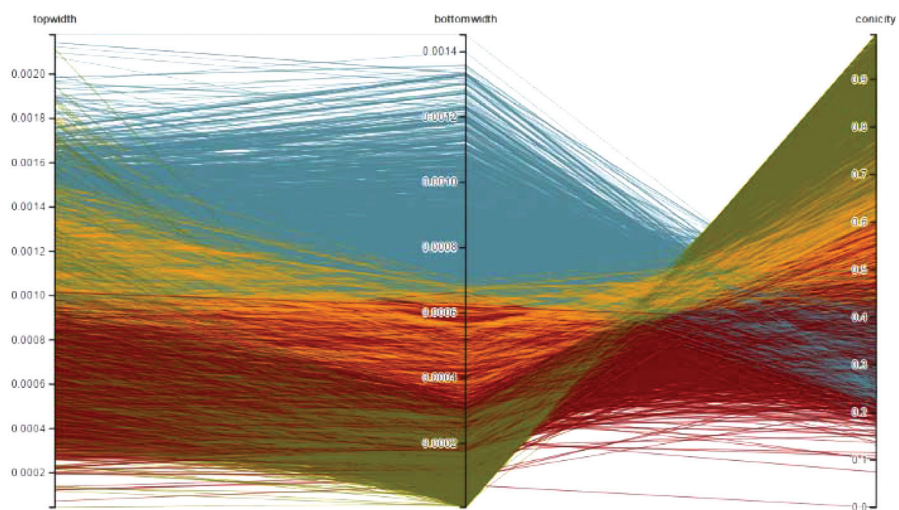
The output of the asymptotic drill reduced model represents the shape of the drilling hole and thus consists of three dimensions: the widths at the top and the bottom of the hole as well as the conicity. In order to depict and analyze this multi-dimensional data, a visualization technique named parallel coordinates is utilized. Figure 4 shows its implementation in the VPI platform for 10,000 laser drilling sampling points, whereas the data is generated with the fast reduce model. In the next step, the data are divided with a clustering algorithm into 4 clusters. The following Figure 5 illustrates the clustering results of the K-means algorithm.

**Table 1** Example of parameters and their range in laser drilling processes

Parameters	Ranges
Pulse Duration [tp]	0.1–1.5 [ms]
Laser Power [PL]	3–10 [kW]
Focal Position [z0]	–8–8 [mm]
Beam Radius [w0]	50–350 [μm]
Rayleigh length [zR]	3–35 [mm]
Workpiece Thickness [d]	0.2–5 [mm]

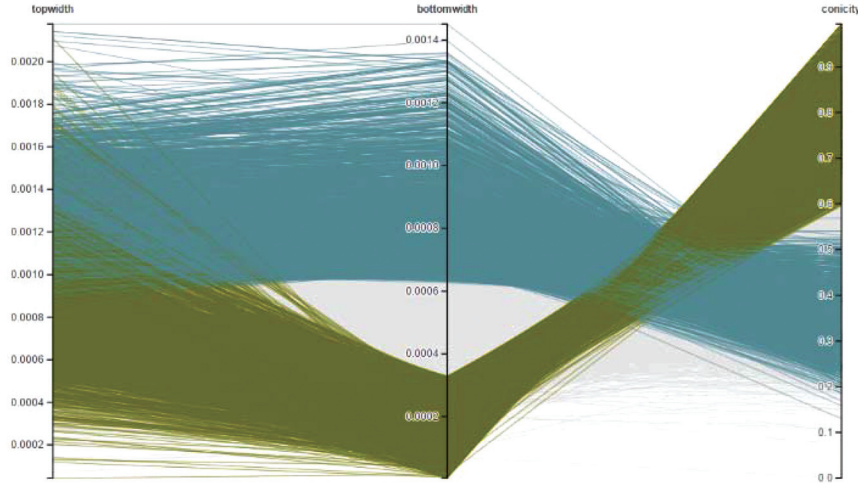


**Figure 4** Parallel coordinates visualization of 10,000 sampling points with 3 different axes (model outputs): top width, bottom width, and conicity.



**Figure 5** Clustering results with four clusters (blue, green, red, yellow) of similar process outputs.





**Figure 6** Highlighting bad process results (blue cluster) and good/desired results (green) cluster in the visualization.

The figure shows that 3-dimensional output space is roughly separable into different groups of sampling points, including the blue cluster with 22% of all data points as well as the yellow (26%), red (31%), and green (21%) clusters.

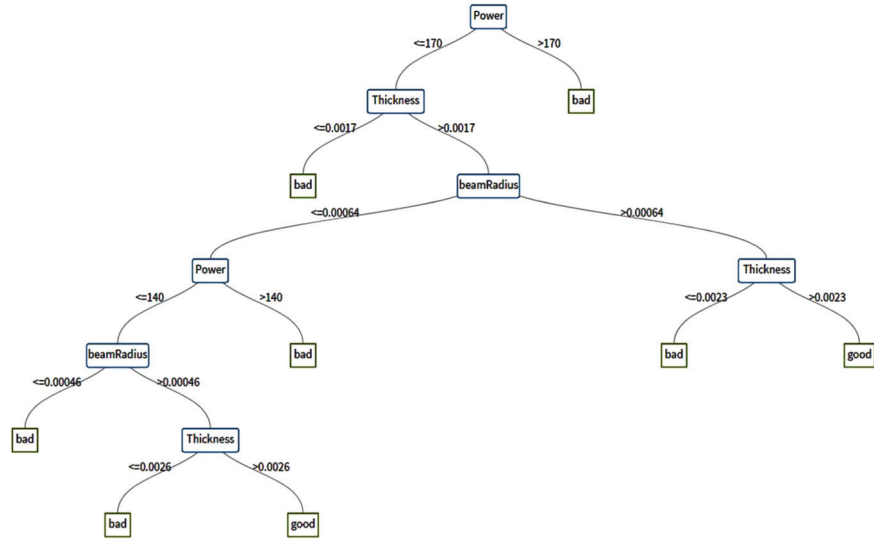
In the following figure, the blue and the green clusters are highlighted. It can be seen that the clusters are lying conversely and that the green one leads to high conicity values. In our application case, the high conicity is a desired process result. Thus, the green cluster is a good one, whereas the other ones are bad.

Having identified the good (i.e. desired) output spaces as well as the bad ones, the next step is to transform the problem into a binary classification problem and to build a classification tree that is used to predict the process outcome (good/bad) on the basis of the laser drilling process parameters (see Table 1).

The following Figure 7 shows a classification tree for the desired clusters (high conicity).

The tree shows that there are mainly two parameter space regions that lead to the desired results (good leaves). These two regions can be defined by the following rules (extracted from the tree):

- *Laser Power*  $\leq 170$  & *Thickness*  $\leq 0.0023$  & *Beam Radius*  $> 0.00064$



**Figure 7** Classification tree that predicts good/bad process outcomes on the basis of the process parameters.

- *Laser Power*  $\leq 140$  & *Thickness*  $> 0.0026$  & *Beam Radius* between 0.00046 and 0.00064

These results show that the hybrid data analytics approach on the top of the reduced model data provides an intuitive and interpretable decision support for the laser drilling process planner. The gained knowledge, especially the identified parameter regions, can subsequently be used to further optimize the process.

## 5 Outlook

In this paper, the methodology to enrich sparse data to dense data and analyze acquired dense data is demonstrated. In order to fully utilize the advantages of the reduced models, a large amount of efforts should be made to generate more useful reduced models. Besides, different reduced models have different domain space, the exploration of the boundary for each reduced model is necessary for generating reliable and high quality process knowledge. These regime conditions can be determined by physical driven or data driven methods.

The comprehensive use of data from reduced models can extract the global and local knowledge of the manufacturing process. The interesting topics can be robustness analysis, as well as the standardization of model reduction process for context specific manufacturing processes.

## **Acknowledgement**

The investigation are partly supported by the German Research Association (DFG) within the Cluster of Excellence “Integrative Production Technology for High-Wage Countries” at RWTH Aachen University.

## **References**

- [1] Brecher, C., *et al.* (2012). Integrative production technology for high-wage countries, Springer, Berlin, Heidelberg.
- [2] Hunter, J. K. (2004). Asymptotic analysis and singular perturbation theory. *Department of Mathematics, University of California at Davis*, 1–3.
- [3] King, J. R., and Riley, D. S. (2000). Asymptotic solutions to the Stefan problem with a constant heat source at the moving boundary. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, The Royal Society, 456(1997), 1163–1174.
- [4] Vossen, G., Hermanns, T., and Schüttler, J. (2013). Analysis and optimal control for free melt flow boundaries in laser cutting with distributed radiation, Wiley Online library, *ZAMMZ. Angew. Math. Mech.*, 1–20/ DOI 10.1002/zamm.201200213.
- [5] Schulz, W., and Michel Freie, J. Randwertaufgaben der thermischen Materialbearbeitung: Inertiale Mannigfaltigkeiten und Dimension des Phasenraums.
- [6] Schulz, W. (1998). *Die Dynamik des thermischen Abtrags mit Grenzschichtcharakter*. PhD Thesis, Aachen: Shaker-Verlag, 2003. Habilitationsschrift, RWTH Aachen.
- [7] Buckingham, E. (1914). On physically similar systems; illustrations of the use of dimensional equations. *Physical review*, 4(4), 345–376.
- [8] Schulz, W., Becker, D., Franke, J., Kemmerling, R., and Herziger, G. (1993). Heat conduction losses in laser cutting of metals. *Journal of Physics D: Applied Physics*, 26(9), 1357–1363.

- [9] Li, Huanrong, Zhendong Luo, and Jing Chen. (2011). “Numerical simulation based on POD for two-dimensional solute transport problems.” *Applied Mathematical Modelling* 35(5), 2489–2498.
- [10] Cordier, L., and Bergmann, M. (2003). Proper orthogonal decomposition: an overview. Post processing of Experimental and numerical Data, Lecture Series 2003/2004, von Karman Institut for Fluid Dynamics, pages 1–45.
- [11] Chatterjee, A. (2000). An introduction to the proper orthogonal decomposition. *Current science*, 78(7), 808–817.
- [12] Hermanns, T., (2018). Interactive process simulation for industrial environments with the example of drilling with laser radiation, PhD thesis.
- [13] Reinhard, R., Büscher, C., Meisen, T., Schilberg, D., and Jeschke, S. (2012). Virtual Production Intelligence – A Contribution to the Digital Factory. In: Hutchison D, Kanade T, Kittler J et al. (eds) *Intelligent Robotics and Applications*, vol. 7506. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 706–715.
- [14] Tercan, H., Al Khawli, T., Eppelt, U., Büscher, C., Meisen, T., and Jeschke, S. (2017). Improving the laser cutting process design by machine learning techniques. *Production Engineering*, 11(2), 195–203.

## Biographies



**You Wang** studied material science at RWTH Aachen University and got his master degree in 2015. He wrote his master thesis with the title “Modelling and Simulation of Glass Heating Process” at Fraunhofer IPT, Aachen. Now he is working as a research associate at the Department of Nonlinear Dynamics of Laser Manufacturing Processes (NLD). His main research interest is the field of model reduction and meta-modelling of laser manufacturing processes. Meta modelling techniques are aiming to set up fast responding and accurate data-driven models by analyzing numeric models or experimental data with multi-dimensional parameters. These performant meta-models lead to fast and frugal customer simulation tools which will strongly support industrial decision making processes.



**Hasan Tercan** has been a scientific researcher at the Cybernetics Lab since July 2015. Mr. Tercan studied computer science at the Technical University of Darmstadt until January 2015. His major fields of study were databases, data warehousing and data analytics. In his master thesis, he investigated the use of various machine learning methods in the financial sector. At the IMA, Mr. Tercan investigates the use of methods of machine learning and artificial intelligence in the production context. The focus here is on AI-supported systems for decision support in production planning as well as automation in production processes.



**Torsten Hermanns** studied physics at RWTH Aachen and received his degree in 2012. His diploma thesis was written at the department “Nonlinear Dynamics of Laser Processing” (NLD) of RWTH Aachen University. In his thesis with the title “Mathematical Modelling and linear Stability Analysis of Laser Fusion Cutting” he derived a stability criterion for the melt film in laser fusion cutting. This stability criterion considered, for the first time, the intensity distribution of the laser beam. After completing his studies, he started working as a research associate at NLD on his dissertation.

Torsten Hermanns is focusing on modelling and simulation of the laser cutting of metallic materials as well as laser processing with short or ultra-short

pulsed laser radiation. This is accomplished by developing reduced models that are based on integral, spectral or asymptotic methods as well as numerical methods. Furthermore, he is responsible for the development of software solutions especially designed for on-site use at the customer.



**Wolfgang Schulz** studied physics at Braunschweig University of Technology. He graduated from the Institute for Theoretical Physics and received a postgraduate scholarship in 1986 on the topic of “Hot electrons in metals”. In 1987, he accepted an invitation to the department Laser Technology at RWTH Aachen University. He received the “Borchers Medal” award in 1992 in recognition of his PhD thesis. In 1997, he joined the Fraunhofer Institute for Laser Technology in Aachen and, in 1999, received the “Venia Legendi” in the field “Principles of Continuum Physics applied to Laser Technology”. His postdoctoral lecture qualification (habilitation) was awarded with the prize of the Friedrich-Wilhelm Foundation at RWTH Aachen University. Since March 2005, he has represented the newly founded department “Nonlinear Dynamics of Laser Processing” at RWTH Aachen University and is the head of the newly founded department of “Modelling and Simulation” at the Fraunhofer Institute for Laser Technology in Aachen. Since 2007, he is the coordinator of the Excellence Cluster Domain “Virtual Production” at RWTH Aachen University.

His current work is focused on developing and improving laser systems and their industrial applications by combination of mathematical, physical and experimental methods. In particular, he applies the principles of optics, continuum physics and thermodynamics to analyse the phenomena involved in laser processing. The mathematical objectives are modelling, analysis and dynamical simulation of Free Boundary Problems, which are systems of nonlinear partial differential equations. Analytical and numerical methods for model reduction are developed and applied. The mathematical analysis yield approximate dynamical systems with small dimension in the phase space and is based on asymptotic properties like the existence of inertial manifolds.