Forschungsberichte aus dem
wbk Institut für Produktionstechnik
Karlsruher Institut für Technologie (KIT)

Leonard Vincent Overbeck

Digital Twins of production systems

Automated validation and update of material flow simulation models with real data

Band 271

Forschungsberichte aus dem
wbk Institut für Produktionstechnik
Karlsruher Institut für Technologie (KIT)

Hrsg.: Prof. Dr.-Ing. Jürgen Fleischer
Prof. Dr.-Ing. Gisela Lanza
Prof. Dr.-Ing. habil. Volker Schulze

Leonard Vincent Overbeck

Digital Twins of production systems
Automated validation and update of material flow simulation models with real data

Band 271

Institut für Produktionstechnik

Digital Twins of production systems
Automated validation and update of material flow simulation models with real data

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

von der KIT-Fakultät für Maschinenbau des

Karlsruher Instituts für Technologie (KIT)

angenommene

Dissertation

von

Leonard Vincent Overbeck, M.Sc.

Tag der mündlichen Prüfung:	14.08.2023
Hauptreferentin:	Prof. Dr.-Ing. Gisela Lanza (KIT)
Korreferent:	Prof. Stephen C. Graves (MIT)

Karlsruher Institut für Technologie

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the internet at http://dnb.d-nb.de.

Zugl.: Karlsruhe, Karlsruher Institut für Technologie, Diss., 2023

Printed in Germany.

ISBN 978-3-8440-9242-4
ISSN 0724-4967

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren
Phone: 0049/2421/99011-0 • Telefax: 0049/2421/99011-9
Internet: www.shaker.de • e-mail: info@shaker.de

Vorwort des Herausgebers

Die schnelle und effiziente Umsetzung innovativer Technologien wird vor dem Hintergrund der Globalisierung der Wirtschaft der entscheidende Wirtschaftsfaktor für produzierende Unternehmen. Universitäten können als "Wertschöpfungspartner" einen wesentlichen Beitrag zur Wettbewerbsfähigkeit der Industrie leisten, indem sie wissenschaftliche Grundlagen sowie neue Methoden und Technologien erarbeiten und aktiv den Umsetzungsprozess in die praktische Anwendung unterstützen.

Vor diesem Hintergrund wird im Rahmen dieser Schriftenreihe über aktuelle Forschungsergebnisse des Instituts für Produktionstechnik (wbk) am Karlsruher Institut für Technologie (KIT) berichtet. Unsere Forschungsarbeiten beschäftigen sich sowohl mit der Leistungssteigerung von additiven und subtraktiven Fertigungsverfahren, den Produktionsanlagen und der Prozessautomatisierung sowie mit der ganzheitlichen Betrachtung und Optimierung der Produktionssysteme und -netzwerke. Hierbei werden jeweils technologische wie auch organisatorische Aspekte betrachtet.

Prof. Dr.-Ing. Jürgen Fleischer

Prof. Dr.-Ing. Gisela Lanza

Prof. Dr.-Ing. habil. Volker Schulze

Preface of the author

I wrote this thesis during my time as research associate at wbk Institute of Production science. Troughout this time, I was supported by many. First of all, I am grateful to my supervisor Prof. Dr.-Ing. Gisela Lanza, whose guidance and help were of great importance for the successful completion of this thesis.

I also have to thank Prof. Stephen C. Graves for hosting me at MIT, many fruitful discussions, and helpful questions which greatly improved the quality of my work. I thank Prof. Dr. Frank Gauterin for chairing my Ph.D. defense in a very professional and pleasant fashion.

This work would not have been possible without the many collaborators at my research partner Bosch. In particular, I want to name and thank Sebastian Brück, Tobias Lechler, Annette Weise, Oliver Romoser, Heiko Schmidt, and Andreas Heupel. My colleagues at wbk, especially in the research group on production systems, were a constant source of motivation and inspiration for me and contributed to this thesis in countless discussions. Representatives for many others, I have to extend a big "Thank you" to Oliver Brützel, Marvin Carl May, Sebastian Behrendt, Andreas Kuhnle, Nicole Stricker, and Martin Benfer.

During my time at wbk, I had the honor to advise many students and their projects impacted my research. Especially, I have to thank Thomas Haizmann, Arthur Le Louarn, Xinyi Xie, Lisa Kudlik, Raphael Kienzler, Christian Merker, Alexander Rose, Michael Teufel, Micha Janikovits, Dominik Wöllstein, and Marius Nagel for their great work, enthusiasm and debates.

I thank the Karlsruhe House of Young Scientists (KHYS) for funding my research stay in Boston, USA.

I would also like to express my sincere gratitude to my parents, who have always supported me completely on my life's journey. They gave me the freedom to learn and the roots to grow. Their example is a powerful motivation for me.

Last but not least, I thank my wife Andréanne for her support in all situations, her advice, and her love. I cannot express what her backing and faith means to me.

Karlsruhe, August 14th, 2023

Leonard Vincent Overbeck

Abstract

To achieve good economic efficiency and sustainability, production systems must be operated at a high level of productivity over long periods. This poses great challenges for manufacturing companies, especially in times of increased volatility, caused, amongst others, by technological transformation in the mobility sector, as well as political and social change, which lead to constantly evolving requirements on the production system. Because the frequency of necessary adaptation decisions and subsequent optimization measures is increasing, the need for evaluation capabilities of scenarios and possible system configurations is growing. A widely applicable powerful tool for this purpose is material flow simulation, but its use is currently limited by its time-consuming manual creation and its limited, project-based usage. A long-term, life cycle accompanying use is currently hindered by the simulation model's labor-intensive maintenance, i.e. the model's manual adaptation in case of changes in the real system.

This thesis aims to develop and implement a concept including the necessary methods to automate the simulation model's maintenance and adaptation to reality and improve the model's accuracy. For this purpose, digital data from the real system are used, which are increasingly available due to trends such as Industry 4.0 and digitalization in general. The pursued vision of this work is a Digital Twin of the production system, which represents a realistic image of the system in the long term through the data-based comparison with reality and its adaptation to reality. This Digital Twin can be used for the realistic evaluation of scenarios, actions, and improvement measures. Therefore, an overall concept and mechanisms for automatic validation and updating of the model were developed. Among other things, the focus was on the development of algorithms for the detection of changes in the structure and processes in the production system, as well as on the study of the influence of the available data on the achievable quality of the Digital Twin.

The developed components could be successfully applied to a real industrial use case at the Robert Bosch GmbH where it lead to a high accuracy Digital Twin, which was successfully used for production planning and improvement. The potential of localization data for the creation of Digital Twins of production systems could be shown in the laboratory environment of the learning factory at the wbk Institute of Production Science.

Kurzfassung

Um eine gute Wirtschaftlichkeit und Nachhaltigkeit zu erzielen müssen Produktionssysteme über lange Zeiträume mit einer hohen Produktivität betrieben werden. Dies stellt produzierende Unternehmen insbesondere in Zeiten gesteigerter Volatilität, die z.B. durch technologische Umbrüche in der Mobilität, sowie politischen und gesellschaftlichen Wandel ausgelöst wird, vor große Herausforderungen, da sich die Anforderungen an das Produktionssystem ständig verändern. Die Frequenz von notwendigen Anpassungsentscheidungen und folgenden Optimierungsmaßnahmen steigt, sodass der Bedarf nach Bewertungsmöglichkeiten von Szenarien und möglichen Systemkonfigurationen zunimmt. Ein mächtiges Werkzeug hierzu ist die Materialflusssimulation, deren Einsatz aktuell jedoch durch ihre aufwändige manuelle Erstellung und ihre zeitlich begrenzte, projektbasierte Nutzung eingeschränkt wird. Einer längerfristigen, lebenszyklusbegleitenden Nutzung steht momentan die arbeitsintensive Pflege des Simulationsmodells, d.h. die manuelle Anpassung des Modells bei Veränderungen am Realsystem, im Wege. Das Ziel der vorliegenden Arbeit ist die Entwicklung und Umsetzung eines Konzeptes inkl. der benötigten Methoden, die Pflege und Anpassung des Simulationsmodells an die Realität zu automatisieren. Hierzu werden die zur Verfügung stehenden Realdaten genutzt, die aufgrund von Trends wie Industrie 4.0 und allgemeiner Digitalisierung verstärkt vorliegen. Die verfolgte Vision der Arbeit ist ein Digitaler Zwilling des Produktionssystems, der durch den Dateninput zu jedem Zeitpunkt ein realitätsnahes Abbild des Systems darstellt und zur realistischen Bewertung von Szenarien verwendet werden kann. Hierfür wurde das benötigte Gesamtkonzept entworfen und die Mechanismen zur automatischen Validierung und Aktualisierung des Modells entwickelt. Im Fokus standen dabei unter anderem die Entwicklung von Algorithmen zur Erkennung von Veränderungen in der Struktur und den Abläufen im Produktionssystem, sowie die Untersuchung des Einflusses der zur Verfügung stehenden Daten.

Die entwickelten Komponenten konnten an einem realen Anwendungsfall der Robert Bosch GmbH erfolgreich eingesetzt werden und führten zu einer Steigerung der Realitätsnähe des Digitalen Zwillings, der erfolgreich zur Produktionsplanung und -optimierung eingesetzt werden konnte. Das Potential von Lokalisierungsdaten für die Erstellung von Digitalen Zwillingen von Produktionssystem konnte anhand der Versuchsumgebung der Lernfabrik des wbk Instituts für Produktionstechnik demonstriert werden.

Contents

Contents **I**

Abbreviations and symbols **V**

1 Introduction **1**

1.1 Motivation 1

1.1.1 Industry perspective 1

1.1.2 Research perspective 2

1.2 Objective 3

1.3 Structure of this work 4

2 Fundamentals **5**

2.1 The Digital Twin – basic definitions 5

2.2 Simulation 8

2.2.1 Definitions 8

2.2.2 Simulation classification 13

2.2.3 Discrete-event simulation 15

2.2.4 Procedure for simulation studies 17

2.2.5 Simulation input data 18

2.2.6 Production systems modeling 20

2.2.7 Verification and validation 23

2.2.8 Material flow simulation in production 26

2.3 Data handling in production 28

2.3.1 Data generation and acquisition 29

2.3.2 Data storage 31

2.3.3 Testing statistical distributions of data 33

2.3.4 The clustering algorithm DBSCAN 35

2.3.5 Process mining 37

3 State of research **40**

3.1 Digital Twins in production research 40

3.1.1 Digital Twin classifications 40

3.1.2 General concepts for the use of Digital Twins 43

3.2 Simulation input data management 45

3.3 Validation and verification of simulation models in production 46

3.4 (Partially) automated simulation model generation and parameterization 47

3.5 Core Manufacturing Simulation Data (CMSD) 52

3.6 Detection of dynamic behavior 53

3.7 Detection of system structure 55

3.8 Research gap 56

4 Own approach 59

4.1 Overall concept 59

4.1.1 Digital Twin procedure 59

4.1.2 Seed model 61

4.2 Model validation 64

4.2.1 Validation process 64

4.2.2 Validation frequency and triggers 65

4.2.3 Validation period 66

4.2.4 Number of validation simulation runs 66

4.2.5 General conditions of validation period 67

4.2.6 Accuracy metrics 70

4.2.7 Limit values for accuracy metrics 76

4.2.8 Validation on multiple validation periods 76

4.3 Model update 77

4.3.1 Definition of update period 78

4.3.2 Determination of update sequence 79

4.3.3 Query data 80

4.3.4 Simulation input computation 81

4.3.5 Update model database 91

5 Prototypical implementation and testing 93

5.1 Use case 93

5.1.1 Production system 93

5.1.2 IT systems 95

5.1.3 Seed model at Bosch use case 96

5.2 Model validation 102

5.2.1 Implementation of validation process 103

5.2.2 Definition of the validation period 104

5.2.3 General conditions of validation period 104

5.2.4 Accuracy metrics and limits 110

5.2.5 Number of validation simulation runs 111

5.3 Model update 112

5.3.1 Definition of update period 112

5.3.2 Definition of update sequence 113

5.3.3 Simulation input computation 113

5.4 Model update with localization data 125

5.4.1 Production system learning factory 125

5.4.2 IT systems learning factory 127

5.4.3 Seed model learning factory 127

5.4.4 Model update in learning factory 127

6 Analysis of the Digital Twin of the production system 132

6.1 Achievable accuracy 132

6.1.1 Initial accuracy before any update 132

6.1.2 Accuracy over time 133

6.1.3 Influence of update period 136

6.2 Sensitivity analysis of the Digital Twin 138

6.2.1 Influence of data quality 139

6.2.2 Importance of individual information components 147

6.3 Use of the Digital Twin in practice 148

6.3.1 Utilization concept 148

6.3.2 Exemplary potentials of a Digital Twin 150

7 Discussion and outlook 157

8 Summary **161**

Own publications **I**

References **IV**

List of Figures **XXXIII**

List of Tables **XXXVIII**

Appendix **XXXIX**

A1 Example of detailed results of sensitivity analysis XXXIX

A2 User interface for validation and update of the Digital Twin XL

A3 Examples of input tables for simulation model in Digital Twin XLII

Abbreviations and symbols

Abbreviation	Description
AD test	Anderson-Darling test
AE	Absolute Error
AGV	Automated Guided Vehicle
ANSI	American National Standards Institute
API	Application Programming Interface
AR	Augmented Reality
ASMG	Automated Simulation Model Generation
BDA	Big Data Analytics
BPMN	Business Process Model and Notation
CAD	Computer-Aided Design
CIRP	Collège International pour la Recherche en Productique - The International Academy for Production Engineering
CMSD	Core Manufacturing Simulation Data
CPPS	Cyber-Physical Production System
CPS	Cyber-Physical System
DBMS	Database Management System
DES	Discrete-Event Simulation
DT	Digital Twin
DWG	"Drawing" - CAD data format
ERP	Enterprise Resource Planning
IIOT	Industrial Internet of Things
IOT	Internet of Things
IT	Information Technology
JPG	Joint Photographic Experts Group
KPI	Key Performance Indicator
KS test	Kolmogorov-Smirnov test
MES	Manufacturing Execution System

ML	Machine Learning
MRP	Manufacturing Resource Planning
MTM	Methods-time Measurement
MTTR	Mean-Time-To-Repair
NASA	National Aeronautics and Space Administration
NIST	National Institute of Standards and Technology
NRMSE	Normalized Root Mean Squared Error
ODBC	Open Database Connectivity
OEE	Overall Equipment Effectiveness
OPC UA	Open Platform Communications Unified Architecture
PDA	Production Data Acquisition
PDF	Portable Document Format
PLC	Programmable Logic Controller
PPC	Production Planning and Control
PPTX	Microsoft Powerpoint Format
QMS	Quality Management System
RDF	Resource Description Framework
RE	Relative Error
RFID	Radio-Frequency Identification
RL	Reinforcement Learning
SCADA	Supervisory Control and Data Acquisition
SDL	Specification and Description Language
SDX	Simulation Data Exchange
SISO	Simulation Interoperability Standards Organization
SQL	Structured Query Language
STEP	Standard for The Exchange of Product Model Data
SysDT	System Digital Twin
SysML	Systems Modeling Language

UML	Unified Modeling Language
UWB	Ultra-wideband
V&V	Verification and Validation
VDI	Verein Deutscher Ingenieure - Association of German Engineers
VR	Virtual Reality
WIP	Work In Progress
XML	Extensible Markup Language

Symbol	Description	Unit	Value Range
$\mathbb{N}_{real}$, $\mathbb{N}_{sim}$	Number of produced parts in reality / in simulation	-	$\mathbb{N}_0$
y_i^{real}, y_i^{sim}	Value of considered KPI at time i in reality / in simulation	-	$\mathbb{R}$
$\bar{y}^{real}$	Average value of the considered KPI in reality (over the whole period)	-	$\mathbb{R}$
H_0, H_1	Hypotheses 0 and 1	-	-
α	Significance level	-	(0,1)
θ	Test criterion for hypothesis test	-	-
$\mathrm{AD}_{\mathrm{emp}}$	Test statistic of the AD test	-	-
$\mathrm{AD}_{\mathrm{krit}}$	Critical value for the test statistic of the AD test at a chosen significance level	-	$\mathbb{R}^+$
X^2	Chi-square hypothesis test	-	$\mathbb{R}^+$
t_f	Upper limit value for the definition of normal machine failure	-	$\mathbb{R}^+$
t_e	Lower limit value for the definition of exceptional events	-	$\mathbb{R}^+$

1 Introduction

1.1 Motivation

The enhancement of common simulation models into Digital Twins of production systems is being pursued by both industry and research for several reasons.

1.1.1 Industry perspective

On the one hand, accelerating technological change coupled with increasing uncertainty in the political and economic environment, as well as growing fluctuations in demand are making it more difficult for companies to plan and operate their production facilities economically over the long term (Choi et al. 2022; Echsler Minguillon & Lanza 2019). Inherent mutability is needed and frequent reconfigurations of the production systems have to be planned. The frequent adjustments of the system require larger, faster, and, most importantly, improved analysis and planning capabilities in production system planning than in conventional, non-agile production systems (Albrecht et al. 2014). On the other hand, new opportunities arise with regard to data availability and processing due to the advancing digitalization, which can relieve production planners of repetitive work through new tools as well as allow them to achieve higher-quality results (Cheng et al. 2018; Kuo & Kusiak 2019).

A key tool for production planning is discrete-event material flow simulation, which, by modeling dynamic and stochastic processes, allows predictions about future scenarios and alternative system configurations (Andreasson et al. 2019 - 2019). A major drawback that currently severely limits the use of material flow simulations in production planning is the high effort required for model creation while the model's time window of use is limited. This is rooted in the standard organization of the simulation usage as a project (Müller-Sommer 2010, p. 7; Onggo et al. 2010; VDI 2014). This project-based creation and usage approach is primarily due to the high manual maintenance effort of models once they are created, especially in rapidly changing production systems.

The hypothesis of this thesis is that maintenance effort can be reduced by integrating existing, digitally available data from corporate IT systems directly into the simulation model. Figure 1-1 illustrates that with conventional simulation use it is possible that throughout the life cycle of a production system (here simplified based on Landherr et al. (2013, p. 164)) several simulation projects are carried out, each with the creation of a new simulation model (e.g., during planning, for optimization after the start of production, and before the introduction of new product variants after a few years). In contrast, the enhancement of the simulation model into a Digital Twin enables the continuous use of the simulation model throughout its entire life cycle. Thus, the added value that a Digital Twin can generate compared to conventional simulation models results from the longer period of use, in addition to its greater accuracy and reduced maintenance effort. Bruckner et al. (2020) present a survey that indicates that simulation is currently mostly used for milestone checks in development and commissioning. The authors predict simulation a great potential in the future, especially in the operating phase of the production system. They expect that in the future simulation models will accompany their counterparts over their entire life cycle and thus become their Digital Twin, for which this dissertation shows a possible approach.

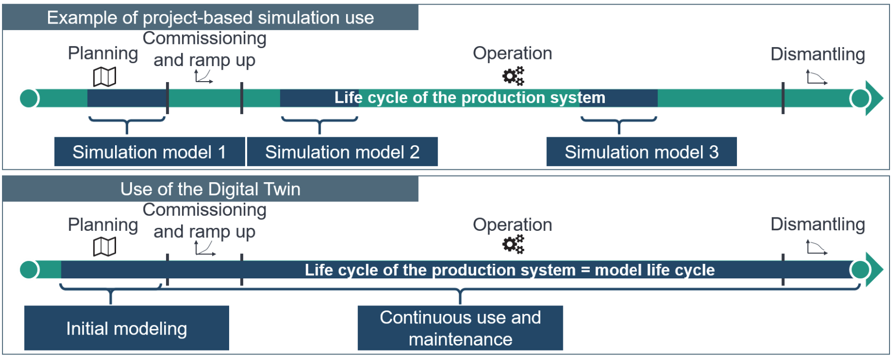


Figure 1-1 Use of conventional simulation models versus the use of a Digital Twin

Current weaknesses that slow down the use of simulation technology in industry today are, according to Manoury et al. (2021), among other things, their project-based use (see Figure 1-1), the necessity to have both distinctive simulation knowledge and expert knowledge of the considered domain, as well as of complex data preparation, which is necessary to achieve a high degree of accuracy. The present dissertation is intended to contribute to the elimination of these drawbacks.

1.1.2 Research perspective

An insight into the historical development, current use, and challenges as well as trends in the field of simulation in manufacturing is provided by Mourtzis (2020). As a key vision for the future, he identifies the transformation towards a Digital Twin that can predict system behavior much more accurately through connectivity and real data (Mourtzis 2020, p. 1941).

Vieira et al. (2018) also provide a research and development agenda in particular for discrete-event simulation (DES) in support of Industry 4.0, specifically addressing visualization capabilities (3D, augmented reality (AR), and virtual reality (VR)) as well as automatic data input and model generation. One of their six identified research gaps is the evolution of DES into Digital Twins (‘R&D Agenda Item #4’), to which the presented work contributes.

1.2 Objective

The goal of this thesis is the development of a generally applicable procedure for the creation of Digital Twins of production systems. The procedure is to be implemented and validated on a use case from industry and in an experimental environment, the learning factory of the wbk Institute of Production Science. Subsequently, the improvement of its accuracy by real data input up to a certain limit shall be examined based on the implemented Digital Twin. Furthermore, a sensitivity analysis of the Digital Twin concerning data availability and quality will be performed, so that statements can be made about the importance of individual data components for the quality of the Digital Twin.

In summary, the objective of this work is **to turn material flow simulation models into Digital Twins of production systems through automated model validation and model update using real data.**

This motivation results in the following requirements for the procedure that shall be developed:

R1 Comprehensive methods and algorithms for data collection and processing to be able to map all relevant aspects of the production system in the Digital Twin. These include:

- R1.1 Model parameters such as process times, availabilities, and scrap rate
- R1.2 Dynamic behavior such as material flow and work flows
- R1.3 System structure such as the number and arrangement of machines and buffers

R2 A coherent and efficient procedure for comparing and adapting the Digital Twin to reality, which must include the following aspects:

- R2.1 Automated validation of the Digital Twin
- R2.2 Automated update of the Digital Twin
- R2.3 Direct applicability to real data

R3 A comprehensive examination of the Digital Twin including:

- R3.1 Implementation in a real use case from industry
- R3.2 Investigation of the resulting behavior of the Digital Twin
- R3.3 Analysis of the influence of the available data on the accuracy

1.3 Structure of this work

The required fundamental knowledge in the areas of simulation, Digital Twins, data in manufacturing companies, and process mining are introduced in chapter 2. The state of the art in research and industry on the topics under discussion is presented in chapter 3. The approach developed in this research is presented in chapter 4. Its application in two use cases is described in chapter 5. The obtained results are presented and discussed in chapter 6. Based on this, the approach and its results are critically evaluated in chapter 7, which also includes an outlook on subsequent research questions. The thesis concludes with a summary in chapter 8.

2 Fundamentals

For the comprehensibility of the presented work, a common understanding of some terms and contexts is indispensable. These include the areas of Digital Twin, which is the vision of this thesis, simulation which lies at the core of the Digital Twin approach, and the capture, storage and processing of data in manufacturing companies which are the indispensable enabler for the approach.

2.1 The Digital Twin – basic definitions

After the concept was first introduced by Michael Grieves in a course on product lifecycle management in 2002 at the University of Michigan (then called the 'Mirrored Spaces Model') (Grieves 2014), NASA (National Aeronautics and Space Administration) revisited the concept in 2012. They defined the Digital Twin as an *"integrated multi-physics, multi-scale, probabilistic simulation of a vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its flying twin."* (Shafto et al. 2010, p. TA11-7).

The term 'Digital Twin' is used today in many contexts and with different definitions and interpretations. A complete presentation and comparison of all existing directions is not possible here for reasons of space and is not necessary for understanding the research project. This dissertation follows the understanding of Barricelli et al. (2019): "*A Digital Twin (DT) is more than a simple model or simulation […]. A DT is a **living, intelligent and evolving model**, being the virtual counterpart of a physical entity or process. It **follows the lifecycle of** its physical twin to monitor, control, and optimize its processes and functions. It continuously **predicts future statuses** (e.g., defects, damages, failures), and allows **simulating and testing novel configurations**, in order to preventively apply maintenance operations."*

This is in agreement with the definition of Liu et al. (2018) *"The Digital Twin is actually a living model of the physical asset or system, which continually adapts to operational changes based on the collected online data and information, and can forecast the future of the corresponding physical counterpart."*

This understanding is also in line with Kuhn (2017) who gives the following definition: *"Digital twins are digital representations of things from the real world. They describe both physical objects and non-physical things such as services by making all relevant information and services available using a uniform interface. For the Digital Twin, it is* *irrelevant whether the counterpart already exists in the real world or will exist in the future. [...]. Often, these are simulation models that simulate, for example, functional or physical properties of the Digital Twin."* According to him, in 2017 the most prominent application area of Digital Twins is production engineering.

The International Academy of Production Engineering CIRP defines Digital Twin in the CIRP Encyclopedia of Production Engineering as follows: "*A Digital Twin is a digital representation of an active unique product (real device, object, machine, service, or intangible asset) or unique product-service system (a system consisting of a product and a related service) that comprises its selected characteristics, properties, conditions, and behaviors by means of models, information, and data within a single or even across multiple life cycle phases."* (Stark & Damerau 2019). This general definition can be understood as the lowest common denominator of many publications, as it does not include the existence of a data link or the ability to predict states.

Rosen et al. (2015) show the historical relation of the Digital Twin concept and present it, as shown in Figure 2-1, as the next stage of simulation technology, in which simulation becomes an integral part of the product-service system and accompanies it throughout its life cycle. This understanding of Digital Twin is consistent with Figure 1-1 which was shown as part of the motivation for this thesis.

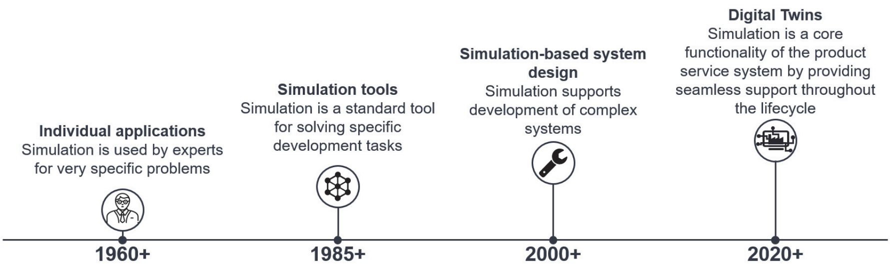


Figure 2-1 Digital Twins as the next evolutionary step in simulation technology entails its integration into the running systems (Rosen et al. 2015)

A frequently used structuring approach of Stark et al. (2017) divides the Digital Twin into a Digital Master, which is a general and abstracted description of a group or class of entities, and the Digital Shadow, which is the collection of all data related to a concrete entity over its entire lifecycle. If the abstract Digital Master is logically linked with the Digital Shadow of a concrete instance, a Digital Twin is created (see Figure 2-2).

Detailed literature reviews on the topic of Digital Twins in production are provided by Negri et al. (2017) and Kritzinger et al. (2018). A compilation of the different characterizations of Digital Twins was created by Jones et al. (2020). The identified research gaps are, amongst others, the consideration of the Digital Twin across the entire system lifecycle and its technical implementation. These open questions will be considered in the research presented here.

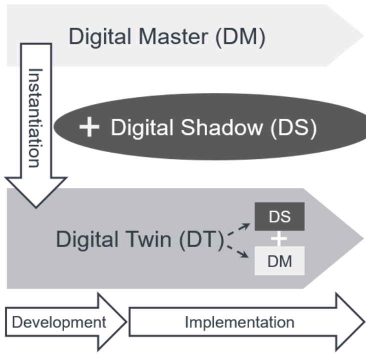


Figure 2-2 Components of the Digital Twin according to Stark et al. (2017)

Building on his initial coinage of the term, Michael Grieves (2014) describes how advances in IT have increased the capabilities of data collection, analysis, and modeling since the initial idea of the Digital Twin so that it can now be broadly applied in the real world. This definition of the Digital Twin includes three components: the physical part, the virtual part, and the data link between the two. His focus remains on the Digital Twin of products. When he talks about the simulation of the production process, he focuses on its impact on the product. After improved data capture in the physical world and more powerful virtual models, he sees pent-up demand for bidirectional data connectivity between the two worlds. Grieves distinguishes possible use cases for Digital Twins according to the ways they help people: conceptualization (through visualization), comparison, and collaboration (through shared discussion).

Grieves & Vickers (2017) deal in greater detail with the theoretical concept of the Digital Twin and its implications especially from a systems theory perspective. According to their paper, the idea of Digital Twins, which has remained the same over all the years in which the terminology has changed, is that all digital information and models relating to a physical system are considered and treated in sum as an independent entity. They distinguish between Digital Twin Prototype and Digital Twin Instance, a classification which has some resemblance to the ideas of Stark et al.

Various standardization activities are already underway for Digital Twins, e.g. by the Industrial Digital Twin Association which uses the Asset Administration Shell as an open-source option[1].

In the context of this dissertation, following the definitions described above, a Digital Twin is understood as a digital image of a physical system that retains a high degree of accuracy over the life cycle of the real system by dynamically adapting to real data and that can predict the behavior of the real system under different scenarios.

2.2 Simulation

Since, as described in section 2.1, the simulation ability is a central property of the Digital Twin, the basics of the subject of simulation, necessary for understanding the developed approach, are summarized below.

Simulations are used in numerous areas and can take on a wide variety of forms, which is why, after some general explanations and classifications on simulation and related terms, the discrete-event material flow simulation that is prevalent for production and logistics systems is presented in greater detail. Afterwards, the organization of simulation usage, the preparation of the required input data will be introduced. Since this thesis treats especially simulation of production systems, their modeling is discussed before multiple aspects of model verification and validation will be elaborated which are also highly relevant for keeping track of the accuracy of the Digital Twin. Because this thesis aims to improve the use of material flow simulation in production, basic insights into this particular subject are given at the end of this section.

2.2.1 Definitions

For the understanding of simulation, the terms *system* and *model* play a central role, which is why these must first be defined before simulation can be explained.

[1] Industrial Digital Twin Association e.V. (2023), The Industrial Digital Twin. industrialdigitaltwin.org [accessed on March 19th, 2023]

2.2.1.1 System

Although a certain range of definitions of the term ‘system’ exists, the three definitions selected here are intended to provide the reader with the understanding underlying this thesis.

The DIN IEC standard 60050-351 describes a system as a "*set of interrelated elements considered in a defined context as a whole and separated from their environment."* (DIN IEC 2004, p. 21). A stronger reference to the function of the system is given by the definition of Grieves & Vickers (2017, p. 87): "*A system is two or more components that combine together to produce from one or more inputs one or more results that could not be obtained from the components individually."*

Gutenschwager et al. (2017, p. 11) list six essential properties that describe a system:

- System boundaries: A system is limited and delimits its scope from the environment by so-called system boundaries. Systems can exchange matter, energy, and information via interfaces at their system boundaries.
- System elements: A system consists of system elements. Up to a certain level of granularity, these can be regarded as systems in their own right (subsystems).
- Structure: The individual system elements are related to each other and thus form a system structure. The structure of a system essentially determines its functionality.
- System state: Each system element has properties that are represented by constant and variable attributes, the so-called state variables. The state of a system element is described by the values of the constant and variable attributes at a certain point in time. The state of a system in turn results from the states of the individual system elements.
- State changes: The states of the system elements can change over time.
- Flow structure: Each system element contains a flow structure, which uses rules to define the state variables and changes.

Complicated systems are characterized by a large number of components that must be coordinated (such as mechanical clocks), while complex systems are characterized by being influenced by external, unpredictable factors (such as the stock market or ecosystems) (Ottino 2003). Because simulation can deal with uncertainty and can be used to evaluate different scenarios it is often used for complex systems, e.g. production systems.

2.2.1.2 Model

One advantage of simulation is that it allows performing experiments not with the system itself but with a model of it. VDI guideline 3633 defines a model as the *"simplified reproduction of a planned or existing system with its processes in a different conceptual or concrete system." (VDI 2014, p. 3).*

Therefore, not all properties of the original system are taken into account during modeling, but those that appear relevant to the modeler are selected (Bracht et al. 2018, p. 83). Models are characterized by a delineation of the system of observation, which consists of interrelated subsystems or elements. The individual elements have properties, perform functions, and are related to other elements. The chosen granularity (or level of abstraction) depends on the modeling goals and is of crucial importance for the statements that can be made with the model. Since a higher level of detail is usually accompanied by higher model complexity, the guiding principle is to model only as detailed as necessary.

Two possibilities for abstraction during modeling are *reduction,* i.e. omitting irrelevant details, and *idealization*, i.e. simplifying system properties. Idealization can be done *inductively*, i.e. by *deriving* individual cases from the totality, or *deductively,* i.e. by *deducing* individual cases from general rules (Gutenschwager et al. 2017, p. 20). In modeling production systems normally all of these possibilities are used in combination.

Models can be assigned to different categories based on the power of the statements can be made with them: Descriptive models describe the system somehow and reflect the system’s interrelationships. Examples of descriptive models are e.g. statistical distributions or visualizations of facts (for example in UML diagramms). Predictive models enable the prognosis of system states and results under defined conditions. These include regression models and simulation models. Prescriptive models can give optimal recommendations for action. For example, mathematical optimization models belong to this class of models. (Lepenioti et al. 2020, p. 58)

2.2.1.3 Simulation

The VDI defines a simulation as the "*representation of a system with its dynamic processes in an experimentable model to reach findings which are transferable to reality; in particular, the processes are developed over time.*" (VDI 2014, p. 3). This largely coincides with the description of Fischer & Hofer (2011, p. 822) of a simulation as *"mapping and imitation of complex, dynamic systems, processes in science, environment, traffic, society, etc. by data processing devices*".

By conducting and evaluating experiments on the simulation model, knowledge can be gained about the behavior of the represented system that can be transferred to reality. With the help of repeatable simulation runs, the influence of systematic parameter variations on the behavior of the model and the target variables in the production system can be investigated experimentally (VDI 2014). Thus, simulation enables the virtual testing and evaluation of scenarios and possible improvement measures and, when applied in production planning, can thus save expensive production shutdowns, as well as frequent test runs and interventions in the real production system (Herbert et al. 2021, p. 139). Other key advantages of simulation are the representation of the dynamic behavior of the system over time, as well as the ability to account for random events (Bracht et al. 2018, p. 117). In addition, simulation models allow the representation of complex systems with a large number of entities and levels (Eley 2012, vii).

The relationships between the terms system, model, and simulation, as well as various simulation options, are shown in Figure 2-3.

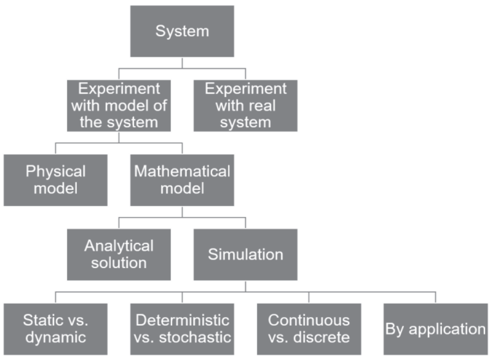


Figure 2-3 Ways to examine a system (based on Law (2015, p. 4))

Queueing models are an alternative modeling, analysis, and optimization option with similar properties, but are based on some assumptions that limit their usability in practice. These include the neglect of transportation times and resources, the non-representability of cyclic material flows, and scrap. Furthermore, solving queueing problems can be computing power expensive, which is frequently resolved by simplifying assumptions. Therefore, simulation is often preferred over queuing models in practice. (Gutenschwager et al. 2017, p. 34)

In production engineering, simulations can be used in the entire life cycle of production systems, i.e. during planning, realization, or operation. Their use is particularly interesting when time- and random-dependent system variables as well as highly interconnected cause-effect relationships are present. In contrast to mathematical optimization methods, simulations are not prescriptive models but only have predictive properties. That said, simulations do not perform mathematical optimization, but support the user by presenting different planning scenarios and observing the respective simulation results so that a good action, but not necessarily the optimal one, can be selected (Banks et al. 2010, p. 23). Thus, the interpretation of the simulation results is ultimately the responsibility of the user, and the simulation only provides targeted support. Also, due to the knowledge required and the time involved, simulation models are currently often only built and operated by simulation experts with skills acquired in specific training (Sitz et al. 2021, p. 147). As a result, the involvement of other departments depends heavily on intensive interaction with the simulation expert. The integration of simulation into the work environment and the company organization could help to improve the manageability and efficiency of working with simulation (Wenzel 2018, p. 30).

If the simulation model is not used actively and creatively by a human to investigate possible scenarios and evaluate actions, as described above, but is instead coupled with any kind of optimization, there are four possibilities according to März (2020):

a. the optimization is integrated with the simulation
b. the simulation is used to evaluate the optimization
c. the simulation determines initial values for the optimization
d. the optimization configures the simulation model

Combinations of these coupling methods result in unlimited possibilities of interaction, which have all one thing in common: They all require a realistic and up-to-date simulation model since the quality of a solution strongly depends on the validity of the simulation model. For this, the presented thesis can help to face this challenge.

If a simulation is used to test real control software, at a time when the system to be controlled does not exist yet, having the simulation model reflect its operation to the controller, it is called emulation (Gutenschwager et al. 2017, pp. 229, 245). This type of methods is not considered in the following.

2.2.2 Simulation classification

Simulations can be categorized according to various properties, the most important of which are shown in Figure 2-3 and are explained below:

Static vs. dynamic
Static simulation models represent the system at a particular point in time, or systems for which time does not matter (e.g., some Monte Carlo models), while dynamic simulation models represent how the system evolves over time (Law 2015, p. 5). This thesis focuses dynamic simulation models.

Deterministic vs. stochastic
In deterministic simulations, the behavior of the system can be predicted a priori with certainty, since no randomness occurs. In stochastic simulations, on the other hand, the occurrence of an event is based on the realization of a random experiment (Law 2015, p. 6). Since a deterministic computer cannot generate true randomness, simulation programs often work with pseudorandom numbers. For this, starting from a given number - the ‘random seed’ -, a series of numbers are calculated that follow the desired probability distribution. Since the numbers are therefore not really random, but only behave as if they were, simulation runs are exactly reproducible when using the same seed value, which increases the traceability and comparability of the simulation results (Eley 2012, p. 24). The presented approach can be used for both types of simulation models but is certainly for stochastic models more interesting since most real complex systems are stochastic in one way or another.

Continuous vs. discrete
This distinction can refer to the set of states as well as to the representation of time. When time is represented as continuous, the time set consists of all positive real numbers including zero; when modeled as discrete, it consists of a countable set of time points that are all equidistant. The states of the system can also be modeled as discrete (e.g., ‘full’ or ‘empty’ as possible states of a tank) or continuous (e.g., filling level of the tank in %). (Gutenschwager et al. 2017, pp. 15–16)

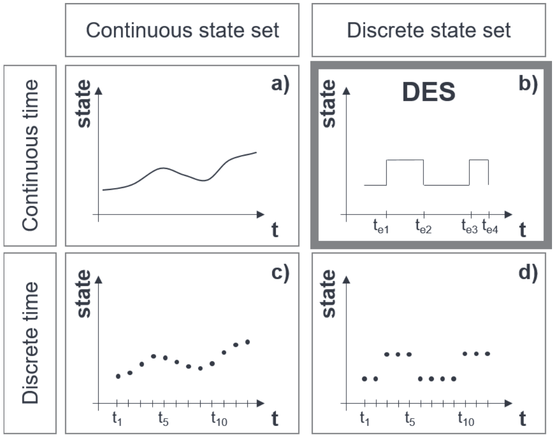


Figure 2-4 Classification of simulation models according to time and state set (Gutenschwager et al. 2017, p. 16) (based on Ören 1979, p. 36)

These distinctions between time and event modeling approaches lead to the modeling alternatives shown in Figure 2-4. The discrete-event simulation (DES) is shown here as case b), i.e. states are discrete and can be switched at arbitrary points in time between them. This type of simulation is particularly common in the mapping of discrete processes (such as discrete production) (Rose & März 2010, p. 14) and is therefore presented in greater detail below and will be the focus of this thesis.

By application

Of course, simulation models can also be differentiated according to their application, whereby the range of possibilities cannot be enumerated completely. For example, in addition to the use in production considered in this thesis, DES alone can also be used among other things for road traffic, flows of people, business processes, or material movements in a warehouse (Hedtstück 2013, p. 117–130). In the presented work, the material flow through a factory or a production system is simulated, so it is a material flow simulation. It can be used, for example, to answer questions on the topics of output, optimization potential, bottlenecks, work leveling, choice of assembly concept, rework strategy, and personnel deployment. (Greinacher et al. 2020; Mayer et al. 2020, p. 129)

2.2.3 Discrete-event simulation

Discrete-event simulation (DES) is a particular type of simulation which will be the focus of this work. Its characteristic is that the progression of time is represented by jumping from event to event. The time progression is thus represented discretely, and the intervals between each time point considered are not of equal length, since the time between events may vary. Typical events in production systems are, for example, the end of a process on a machine, the arrival of a part at the end of a conveyor belt, or the start of a machine failure. Each of these events triggers a change in the state of the system. For the correct representation of the time sequence of all events, the simulation program keeps a list of all upcoming events chronologically. After it has processed one event, the current time is set to the time of the next event on the list and the associated change of state is executed. If this results in further future events, these must be entered at the correct positions in the chronological list. For example, if the current event is the start of a machining process, the calculated end of the process must be added to the list. Then the simulation program moves to the next event in the list. This happens until either the list is empty or a predefined stop criterion has been reached (e.g. set simulation time has been reached). (Gutenschwager et al. 2017, p. 55)

To build a DES model, according to Eley (2012, pp. 9–10) seven (tangible and intangible) basic elements are required in addition to the event list:

- Entities: Individually identifiable objects that can change and move during the simulation run. Examples: orders, products, transport containers
- Resources: Objects present over the entire simulation run that can be claimed by entities. Examples: Machines, assembly stations, workers
- Queues: Sub-type of resources. Examples: warehouse
- Attributes: Properties of entities and resources that define representation and behavior. Examples: occupancy, machine state
- Methods: Procedures in programming languages to control the simulation process
- Variables: Used to store information. Examples: numerical values, tables, lists
- Random numbers: Represent stochastic processes such as failure occurrence or unknown customer demand.

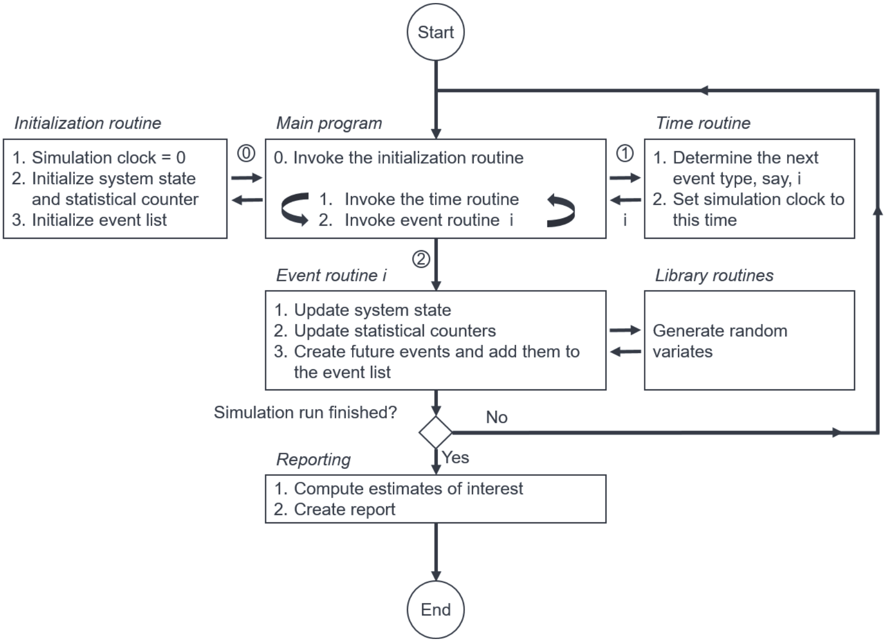


Figure 2-5 Sequence of a DES program (Law 2015, p.10)

Figure 2-5 shows a program flow logic for DES following Law (2015, p. 10). The initialization routine is called first and only once per simulation run. It sets the basic configuration of the model. It resets the simulation clock and creates the initial event list, but also assigns variables, and generates entities and resources. A challenge is to define this initial state according to the model restrictions. Not every start state of the model leads to an error-free model run. Incorrectly initialized start assignments can lead to blockades in the model flow. The main program then calls the timing routine and the event routine alternately. The timing routine determines the next event and sets the simulation time to the time of this event. The event routine updates the system state, updates the statistics counters, generates future events (for which it uses random variables), and adds them to the event list. When the simulation run is completed, the desired reports are generated.

2.2.4 Procedure for simulation studies

As mentioned in the motivation, simulation models are commonly created and used in projects. The VDI guideline 3633 recommends the work flow described in Figure 2-6 for simulation projects.

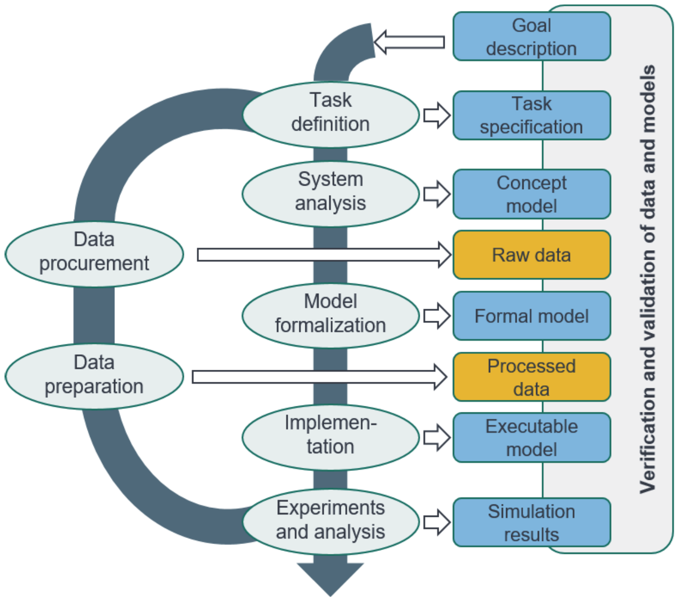


Figure 2-6 Procedure model for simulation projects according to VDI (2014)

Every simulation project should begin with a description of the goal, from which the task definition is derived. One particularity of this guideline is the parallelism of data processing (divided into data acquisition and data preparation) and working on the model (divided into system analysis, model formalization, and implementation). The two strands do not converge again until the end of the project in the experimentation and analysis phase. In addition, after each phase, the documents that should be completed are defined. The continuous monitoring of the simulation project through continuous verification and validation is emphasized.

In addition to this guideline, which is particularly widespread in Germany, numerous other procedural models for simulation projects can be found, of which Santos et al. (2022) summarize some. The approaches are all similar in the logical sequence of steps but differ in naming and granularity. For example, Law (2009) suggests a seven-step approach, while Sargent (2013) presents a procedure with three main steps: problem definition, conceptual modeling, and computational modeling. Greenwood (2020) from FlexSim, a commercial simulation software, proposed a method for simulation project management which covers the five phases of the project initialization, planning, execution, monitoring, and closure. All these approaches assume no changes or updates of the simulation model and a limited time for its usage. As explained in the motivation chapter, the goal of this thesis is to change the way simulation is used.

2.2.5 Simulation input data

Data are the basis for every simulation model and determines the quality and reliability of the statements that can be made with the simulation model (often summarized as ‘garbage in, garbage out’). The data required for material flow simulation models of production systems are often located in a variety of different IT systems (see section 2.3.1) and are available in different formats so the collection and preparation of the data for the simulation model generation are generally very time-consuming (Mieschner & Mayer 2020, p. 319).

Collisi (2002, pp. 21-35) distinguishes between five classes of data concerning simulation models:

- Data class S: describe the structure of the model
- Data class A: describe processes of the model
- Data class P: describe the parameters of the model
- Data class D: used to perform experiments
- Data class E: results of the simulation runs

For Digital Twins of production systems data classes S, A and P are of particular importance for the model update, while D and E are primarily relevant for the model validation.

In addition, Robinson & Bhatia (1995, p. 63) distinguish simulation data according to availability and collectability:

- Category A: Available
- Category B: Not available but collectible
- Category C: Neither available nor collectible

Acél (1992) identifies data collection as the by far most time-consuming activity in simulation model building for production and logistics, accounting for 12-50% of the total work time (out of 13 activities considered).

When estimating the effort required for intralogistics simulations in the automotive industry, data collection and plausibility checks are also rated as the most time-consuming activities (42% of the total effort on average) by Müller-Sommer (2010, pp. 7-9). In particular, the effort/quality ratio of the process data is rated negatively. Skoogh & Johansson (2007) state that on average 31% of project time is invested in input data management. Onggo et al. (2010) estimate the effort for input data management to be 10-40% of the total project time, with an even larger percentage in many projects.

If data is interpreted and given meaning, information can be extracted from it (Hildebrand et al. 2018, p. 5). Important for the general evaluation of the usefulness of the information and thus the basis for weighing the effort that may need to be invested in its acquisition are according to Schmidt-Volkmar (2008, pp. 6–7) in particular:

- Decision relevance: providing decision-makers with a relevant knowledge advantage
- Relevance to time: topicality of the content
- Information content: degree of correspondence of the information with reality, consistency, and quality of the data
- Information preparation: increasing the understanding of the information through integration, aggregation, and structuring.

Robertson & Perera (2002) describe four different methods to connect a simulation model to data. All four approaches are shown in Figure 2-7. In method A, the modeler manually enters all data into the simulation model. In method B, the modeler collects the data manually and then it is stored centrally in files. The simulation then reads these files in. Methods C and D retrieve the simulation data directly from the databases of a company. In method C, the data is extracted from the company databases and stored in an intermediate database, whereas in D all data from the company databases are directly transferred to the model. Robertson & Perera point out that manual data collection is very time-consuming. Therefore, they see the future of data integration in methods C and D that involve a link to the organization's databases. This reduces the manual effort of data acquisition and increases data accuracy and reliability. However, challenges are to ensure the quality of the data in the databases of the company, as well as to establish the connection between the systems. The approach presented in this work uses methods C and D.

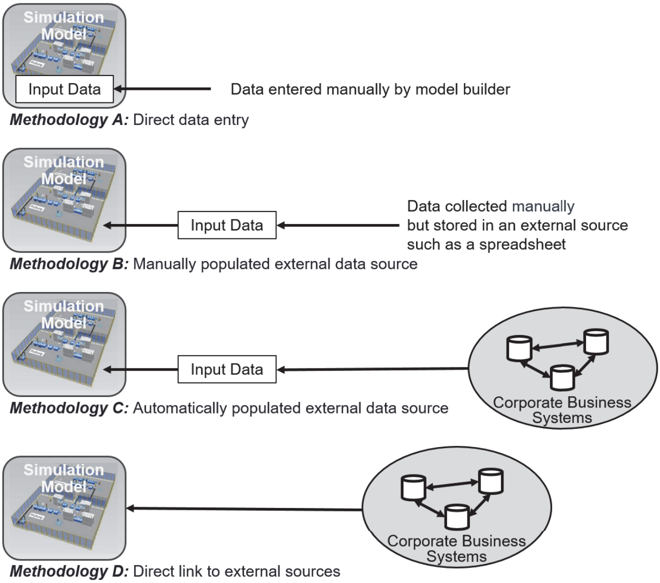


Figure 2-7 Ways to enter data into the simulation model (Robertson & Perera 2002)

2.2.6 Production systems modeling

This dissertation focuses on Digital Twins of production systems. A production system is defined according to Kellner et al. (2020) as a closed set of system elements such as factory halls, workers, machines, conveyors, and storage bins, which are interrelated and jointly produce certain goods. This can be, for example, a workshop area, a production line, a factory floor, or a whole plant. As described in section 2.1 and shown in Figure 2-8, Digital Twins can be created on the different structuring levels of production, but in this work, Digital Twins of the levels of *factory* and *lines* or *production areas* are considered.

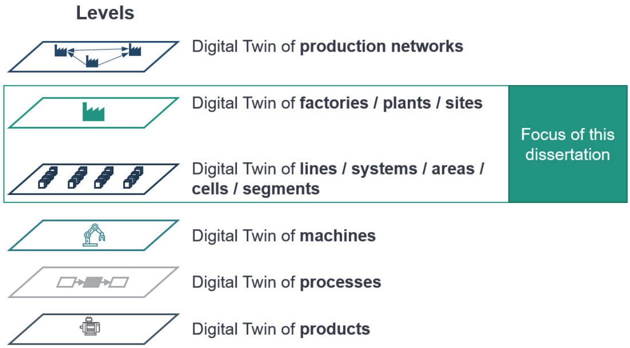


Figure 2-8 Possible Digital Twins on the structuring levels of production, based on (Westkämper 2007, p. 11; Wiendahl et al. 2007)

For the simulation of a production system, first, a model of it has to be created. Although a variety of different modeling concepts exists and combinations of these are possible, the building block-oriented approach is particularly important for the modeling of production and logistics systems in commercial simulation software. They *"... provide predefined model elements (building blocks) in building block libraries for an application domain, with the help of which a model can be built*." (Gutenschwager et al. 2017, p. 16). When defining the predefined building blocks, different approaches can be chosen and also combined. For example, building blocks can be either technology- or process-oriented, permanent or temporary, physical or logical, stationary or mobile. The approach presented in this thesis includes building blocks with different characteristic, i.e. physical machines and logical methods, permanent workers and temporary products. It is important that model instances of building blocks always have a defined state and can be parameterized and combined. Frequently, building blocks also have an internal flow logic that influences the model dynamics.

2.2.6.1 Modeling process

During model creation, the modeler must make numerous decisions about what should be included in the model and what should not be included, as well as which level of detail or abstraction is selected for individual areas. Thus, he selects one of the infinite possible model variants for the respective use case. This is also the case for the creation of simulation models (Robinson et al. 2010, vii). In simulation model building, the Principles of Proper Modeling described by Becker et al. (1995) should be considered:

- *Correctness:* both syntactic (complete and consistent adherence to the chosen modeling rules; independent of the object) and semantic (structural and behavioral fidelity of the model to the real system) correctness
- *Relevance*: the elements and relationships contained in the model increase the usefulness of the model
- *Cost-effectiveness:* creation effort is economically justified
- *Clarity:* addressee-specific; includes aspects of structure, clearness, and readability
- *Comparability:* both syntactically (compatibility of models created with different methods) and semantically (content)
- *Systematic structure:* enables composition of individually modeled components into a comprehensible overall architecture, with the defined modeling rules providing the structuring framework.

2.2.6.2 Process modeling

Modeling of dynamic processes is particularly important and challenging for complex systems as production systems. According to Bichler et al. (2017, p. 180), a process is the "*sequence of operational tasks, each of which represents a sub-process. [. . .] A (sub-) process transforms an input variable [. . .] into an output variable [. . .]. Processes have a clearly defined starting and ending point."*

According to Becker (2018), processes are defined as operations whose results can be clearly described and also show how these results are obtained. The result can be either something tangible like a workpiece or something intangible like information. Processes can be hierarchical so that the main process consists of several sub-processes. The sub-processes can also be subdivided into finer parts, the lowest level is referred to as *activities.* The hierarchical structure of a process is shown in Figure 2-9.

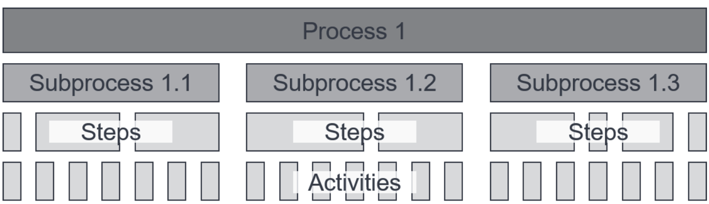


Figure 2-9 Hierarchical structure of a process based on Becker (2018, p. 9)

When breaking down a process into sub-processes and further down to elementary activities, it becomes clear that a process model can be modeled at different levels of detail. Again, the guideline applies that processes should be modeled *"as abstract as possible and as detailed as necessary*" (Hrdliczka et al. 1997, p. 7).

One possibility to describe process flows formally are Petri nets. They were presented for the first time by Carl Adam Petri in his dissertation in 1962 (Petri 1962). These nets are marked and directed graphs consisting of places and transitions. Each transition is preceded and followed by at least one place, which can be marked. A transition can be executed as soon as all directly upstream places are marked. After the execution, also called 'firing', the downstream places are marked, and the marks thus move on. In this process, the markers are referred to as tokens. (Priese & Wimmel 2008, pp. 49-53)

In Figure 2-10, an exemplary Petri net including the central elements' places, transitions, and tokens can be seen. In the state shown, both t1 and t4 can be executed, since all previous places (in this case one each) are occupied by tokens. After the two transitions, s2 and s4 are occupied and thus t3 can be executed, which in turn would lead to the final state s5.

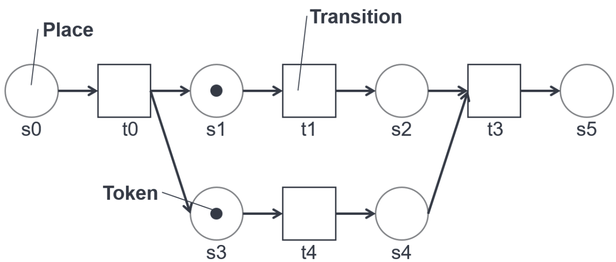


Figure 2-10 Exemplary Petri net and its elements

Because of their general applicability, Petri nets will be one key element for the recognition and description of processes in the approach developed in this thesis. In particular, they are the result of several process mining algorithms (see section 2.3.5).

2.2.7 Verification and validation

A central activity in the creation of simulation models, especially when the focus lies on high accuracy, is verification and validation (V&V). The objective is to check the credibility of the model and to gain the confidence of the decision-makers in the model. Verification is about answering the question ‘Is the model correct?’, while validation is about ‘Is it the right model (concerning the task)?’, i.e. in particular whether the model represents the behavior of the system sufficiently accurately (Gutenschwager et al. 2017, p. 203). As can be seen in Figure 2-6, these questions should accompany the entire course of the simulation project. Aspects of credibility include the feasibility of the simulation study, the correctness of the phase results, and the appropriateness of the results for the application (Rabe et al. 2008, pp. 19–23). This can be achieved with the help of nine criteria: completeness, consistency, accuracy, timeliness, suitability, plausibility, understandability, feasibility, and availability (Rabe et al. 2008, pp. 22–23). These criteria strongly resemble the dimensions in which data quality is measured: completeness, consistency, accuracy, timeliness, and relevance (Günther et al. 2019, p. 585).

A variety of different V&V techniques for simulation models exist, such as structured walk-through, testing of submodels, sensitivity analysis, animation, and comparison with recorded data (Gutenschwager et al. 2017, p. 214).

Five quality criteria can be defined for simulation projects in production and logistics as a whole (Wenzel et al. 2008, p. 5):

1. Careful project preparation
2. Consistent documentation
3. End-to-end verification and validation
4. Continuous integration of the client
5. Systematic project implementation

For simulation models that are not implemented as a project (which would include Digital Twins) additional or different quality criteria may apply, according to the authors.

A possible validation technique for simulation models is the linear regression analysis between simulated and real data, where besides the correlation coefficient also the slope and the regression constant (y-axis intercept) of the regression provide important information about the alignment between the model and reality (Analla 1998).

Concrete instructions for the creation of valid and credible simulation models are according to Law (2019): precise formulation of the task, involvement of domain experts, regular exchange with decision-makers, and the use of quantitative validation techniques at the model component level. Further key actions for valid and credible simulation models are documentation and structured discussion of the assumptions made, sensitivity analysis of the important model parameters, validation of the model output with real output (including the use of statistical methods), and model animation. He enumerates common problems related to data needed to build models (Law 2019, p. 1412):

- Existing data are not representative of what is actually needed for the simulation model
- Data are in an unsuitable format
- Data contain measuring, recording, or rounding errors
- Data are deliberately falsified
- Circumstances of data generation are unclear/unknown

Other aspects besides the purely technical V&V that influence the confidence in simulation models and their results investigate Harper et al. (2021). They go beyond the already discussed effect of the general credibility of the performing simulation expert and discuss, based on extensive existing literature, the interaction between model creator, model, and stakeholder/client in the different phases of a simulation project. This aspect will be revisited in the end of this thesis when a user concept for the Digital Twin is developed.

In a literature review of 61 relevant publications, Kleijnen (1999) examined V&V techniques in simulation models for operations research. For verification, four areas are considered: 1) general programming rules, 2) verification of intermediate results of the model and its modules, 3) statistical testing of final simulation results compared to analytical results, and 4) animation. The focus for model validation is on 1) use of real data, 2) comparison of simulation and real data, 3) correlation analysis of simulation and real result data and comparison of mean values, 4) sensitivity analysis, and 5) white- versus black-box simulation models. In white-box approaches, the interior of the model is visible and comprehensible to the user, whereas this is not possible in black-box approaches.

Model validation techniques can be divided into subjective and objective approaches (Balci 1989). However, the term ‘objective’ refers only to the method itself. The selection of techniques and the decision about the scope of application always contain subjective components (Rabe et al. 2008, p. 115).

The choice of validation techniques is influenced significantly by the observability of the real system. A system is observable if data on system behavior can be collected and used as reference values (Sargent 2020).

Table 2-1 contains a list of possible techniques for operational validation, classified according to the criteria described above.

Table 2-1 Classification of validation techniques with examples (Sargent 2020)

	Observable system	Unobservable system
Subjective approach	• Graphic comparison • Analysis model of behavior	• Analysis of model behavior • Comparison with other models
Objective approach	• Comparison using statistical tests and statistical techniques	• Use of statistical tests for comparison with other models

Output variables can be evaluated with statistical techniques and comparative values from a real system. Even with quite simple approaches, additional credibility can be gained. However, only the validity of the model for the existing system is proven, and by no means for variants to be investigated or redesigns of the system. For simple statistical procedures as well as for examples that are more complex it is true that results from the simulation are time-series, while many statistical techniques assume independent samples. Therefore, a variety of statistical techniques can only be applied after the use of additional procedures for deriving independent sample values from the time series values of the simulation. (Rabe et al. 2008, p. 104)

For the approach for Digital Twins of production systems presented in this thesis especially the validation plays an important role. To make the approach broadly applicable, several objective techniques will be used that work on observable and (partly-) unobservable systems (simulation model vs. real system). Because the output values of simulation and reality normally do not meet the requirements for most statistical tests, other comparisons of statistical values will be used for validation.

2.2.8 Material flow simulation in production

As mentioned above, the core of the Digital Twin of the production system includes a DES material flow simulation model of the system. Material flow simulation is a well-established and intensively studied subject in production. Important questions in the context of Digital Twins is when to use simulation and which simulation tools to use.

Simulation is generally used when (Eley 2012, vii):

- The system to be modeled is extensive, involving many different decision areas of a company or supply chain, and the corresponding mathematical model would become so complex that it would be almost impossible to solve.
- The model is characterized by many stochastic influencing variables that analytical models cannot represent.
- Different solution ideas are already available, but testing them on the real system is impossible, expensive, or risky.
- The user has little experience in applying mathematical optimization methods.

The first three points are true for nearly all production systems, point number four does also apply regularly in practice.

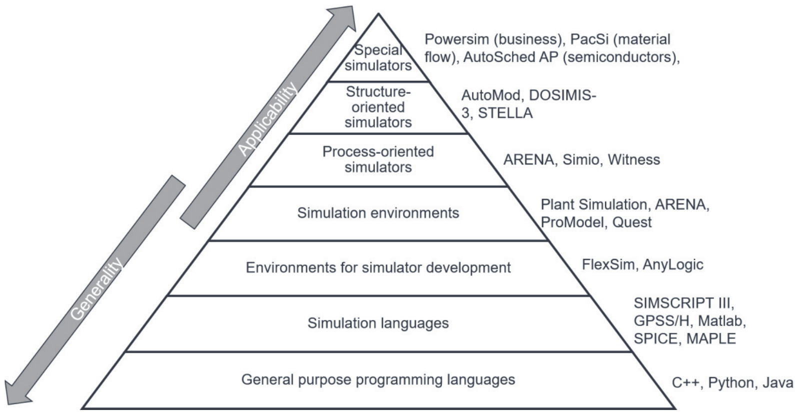


Figure 2-11 Classification of simulation tools for production and logistics, actualized and based on Eley (2012, p. 10) and Wenzel (2008)

So-called simulation tools support the creation of simulation models with computers. As shown in Figure 2-11, these can be sorted according to their application reference or generality. Special-purpose simulators, which are specifically tailored to one specific simulation task, are one extreme case and general-purpose programming languages, which can be used for a wide range of tasks and thus also for simulation tasks, the other (Eley 2012, p. 10). Turing-complete programming languages like C++, Python, or Java, can indeed solve any task[2]. It should be noted that with the generality of the chosen

[2] OpenGenus IQ (2023), non turing complete programming languages. https://iq.opengenus.org/non-turing-complete-programming-languages/ [accessed on March, 19th 2023]

tool, the necessary implementation effort for each use increases. To achieve an appropriate balance between generality and implementation effort, tools from the middle of the pyramid are often suitable, e.g. simulation environments, as for example Plant Simulation, which are also frequently used in practice.

The application of simulation models in production can be roughly divided into the categories of ‘design and planning’ and ‘operation of production systems’ (Negahban & Smith 2014). The benefits of simulation lie in particular in identifying wrong decisions in advance and reducing the planning risk, objectifying the technical discussion through quantitative results, and gaining an understanding of the overall system. This benefit is often difficult to quantify, but there is much evidence and experience that it leads to large savings when the tool is used correctly. (Hrdliczka et al. 1997, p. 30)

Flores-Garcia et al. (2018) consider the challenges associated with the use of simulation in the early stages of production system development when many fundamental changes are still being made, using three example projects carried out in a company over three years. They look at model conceptualization, model implementation, and model usage.

A review on simulation use in production planning and control (PPC) by Jeon & Kim (2016) shows that discrete-event simulation was used in 45% of the 131 publications considered. The most common use cases are procurement management, shop floor scheduling, and process design and planning. The main focus of the use of DES is on shop floor scheduling and production and process design, as for example by Schwarz et al. (1978), who use simulation to evaluate different scheduling policies for automated warehousing systems, or Greinacher et al. (2020) who use DES to evaluate different production concepts with various measures of the lean production philosophy or measures for higher resource efficiency.

An international ranking of popular discrete event simulation software by Dias et al. (2016) based on 50 parameters is led by Arena, ProModel, and FlexSim. The programs Plant Simulation and AnyLogic, which are particularly widespread in Germany, come in 8th and 9th place. Plant Simulation is used in this thesis.

2.3 Data handling in production

In modern companies - and there especially in production - large amounts of data are generated. Since data form the basis for Digital Twins of the production system, their creation, acquisition, storage, and processing must be precisely understood. Three aspects of data processing will be explained in particular because they are applied in this thesis: tests to identify the statistical distribution of data, clustering algorithms (i.e. DBSCAN), and process mining.

2.3.1 Data generation and acquisition

Around the production of products, companies generate much of the data necessary for simulation model creation and maintenance. These data can be generated at different levels of detail and stored in different IT systems depending on the circumstances of generation and application. Therefore, an overview of the vertical data structure of a manufacturing company is given first. This ranges from physical data collection at the production process to highly aggregated overviews for management and other special purpose programs.

2.3.1.1 The automation pyramid

The IT systems relevant to production are often arranged hierarchically in the so-called automation pyramid in Figure 2-12. There are different versions for the number, naming, and contents of the individual levels (Meudt et al. 2017). Due to its widespread use, this dissertation will refer to the definition by Siepmann (2016).

For a good understanding of data creation, handling, and processing, each level of this hierarchy is important.

2.3.1.1.1 Actuators and sensors

The actuators and sensors that convert electronic signals into physical processes or physically measured values into electronic signals link the digital world and the physical production process. Examples of actuators are rotating and linear drives, pneumatic, hydraulic, or piezoelectric actuators. (Roddeck 2019, pp. 161, 214)

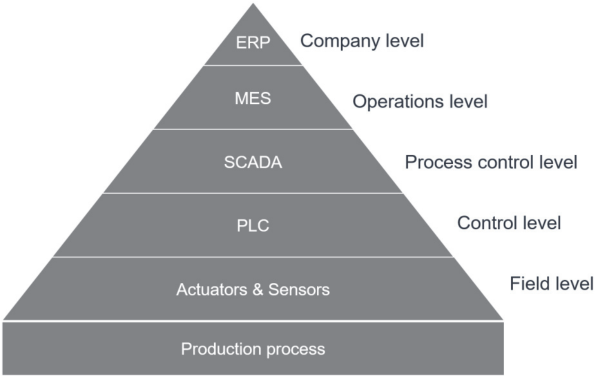


Figure 2-12 The automation pyramid (Siepmann 2016)

2.3.1.1.2 PLC

The processes in individual machines, robots, and plants are programmed in programmable logic controllers (PLCs). These receive signals from sensors (e.g. light barriers, proximity switches, thermometers) from the field level and derive actions from them that are executed by actuators. The PLC processes the program steps at a fixed rate so that the duration of the program execution is precisely predictable and can therefore be referred to as real-time. PLCs are usually located near the controlled equipment in control cabinets on the factory floor. (Wellenreuther & Zastrow 2015)

Further, application-specific features of the PLC are: particularly resistant device components, a simple programming language, permanent operation, signal processing on bit level, and an operating system in the firmware (Fischer & Hofer 2011, p. 795).

2.3.1.1.3 SCADA

Supervisory Control and Data Acquisition (SCADA) refers to a computer system for monitoring and controlling technical processes. They are often used in large industrial plants such as in the process industry, refineries, or power plants (Ghosh & Sampalli 2019). At this level of the automation pyramid, so-called operational data acquisition takes place, which is then passed on to the next higher layer, the MES.

2.3.1.1.4 MES

Manufacturing Execution Systems (MES) are central to the collection, processing, and storage of data at the production system level. Here, data from sensors, PLCs, and

SCADA are aggregated and used to plan and control production processes in real-time. They are also used to create transparency in production. (Kletti & Deisenroth 2021; VDI 2016)

2.3.1.1.5 ERP

Enterprise Resource Planning (ERP) systems are used to structure all resources available in a company and thus contain all master data of the company. In addition to a variety of operational functions, the ERP system covers resource allocation in particular. This function is often also referred to as manufacturing resource planning (MRP) (Kurbel 2013, p. 2). Hence, the distribution of tasks between MES and MRP might vary between companies and sites. The stored data includes information about workpieces, as well as the allocation of all necessary resources for the production of a workpiece. This includes parts lists with all the individual components that make up a part, as well as information about the required machines and workers (Shehab et al. 2004).

2.3.2 Data storage

For the Digital Twin of the production system, the use of real data is particularly important, which is why central aspects of the storage of these data and, in particular, the concept of the data lake will be briefly discussed.

2.3.2.1 Databases

According to (Steiner 2021, p. 6) a database is *"an independent and permanently designed data organization that can manage a data stock securely and flexibly."* Databases are managed with a database management system (DBMS). Various concepts exist for sorting the stored data, the most important distinction being between relational and non-relational database structures.

In general, data can be distinguished according to the degree to which it is structured. Structured data have a predefined and known structure and are usually available in relational databases. Semi-structured data follow a self-describing structure as is the case in CSV or XML files, but lack the formally defined structure of a relational database. Unstructured data does not follow a structure, as is the case with raw text or photographs, for example. (Salam & Stevens 2006, pp. 11-12)

2.3.2.1.1 Relational databases

In relational databases, data are not stored hierarchically in a single file, but in several tables sorted by topics (so-called entities). Storage is thus flexible, as more data can be included and new relationships between data can be easily added. Disadvantages are the decreasing clarity with a large number of tables, their rigid structure, as well as relatively slow queries since the information from different tables has to be merged. (Steiner 2021, p. 9)

Queries to relational databases are performed with special languages, such as SQL (Structured Query Language), the best-known query language published by IBM in 1976 and standardized by the American National Standards Institute (ANSI) in 1986 (Fischer & Hofer 2011, p. 796). One way programs can communicate with relational databases is the ODBC (Open Database Connectivity) specification published in 1992 by Microsoft (Fischer & Hofer 2011, p. 584). These technologies will be used for the Digital Twin approach of this thesis.

2.3.2.1.2 Non-relational databases

Non-relational databases incorporate all systems that do not store data in distributed tables. These include hierarchical or network models, which were common before the advent of relational databases and are still used in some specialized applications, such as CAD (computer-aided design) programs. With the rise of the internet, web applications, and Big Data, non-relational database systems have taken off again. (Meier & Kaufmann 2016, p. 18)

One type of non-relational databases are graph databases, which use graphs to store highly interconnected data. A popular example is the Resource Description Framework (RDF). Special query language, as for example SPARQL, can be used to retrieve the related data in short and constant time, because less joining of different tables is required (Angles et al. 2018).

2.3.2.1.3 Data lake

A popular concept for storing large amounts of structured and unstructured data is the data lake, whose implementation is closely linked to the open-source software framework Apache Hadoop. This enables complex computational operations to be performed on large data sets in computer clusters (Nandimath et al. 2013; Vavilapalli et al. 2013).

It can do so using low-cost and easily accessible technologies for hardware and software. Data lakes act as enterprise-wide data management platforms to make disparate data sources available in their original format. Among other things, this can avoid prior data transformation steps and the simplified accessibility is expected to increase agility in data analytics (Fang 2015). In this context, the data lake concept is replacing the traditional data warehouse for Big Data applications, whose differences are enumerated in Table 2-2 (Khine & Wang 2018). Schema-on-write means that data must be structured already when it is stored, whereas schema-on-read means that data are structured only when they are retrieved.

Table 2-2 Comparison of data warehouse and data lake (Khine & Wang 2018, p. 6)

	data warehouse	**data lake**
data	structured, prepared	structured / semi-structured / unstructured / raw
preparation	schema-on-write	schema-on-read
memory	expensive, reliable	cheap memory
agility	fixed configuration	flexible configuration
users	IT professionals	data scientist

The Digital Twin approach of this thesis will be applied to an industrial use case, which uses a data lake for its production data.

All of the different kinds of databases can be hosted either on server(s) of the company (called on-premise) or on server(s) of another company which are typically part of big computer cluster and accessed over the internet (called cloud). By using the cloud, companies do not have to invest in the hardware themselves but pay a more flexible fee, i.e. pay-per-use. (Repschläger et al. 2010)

2.3.3 Testing statistical distributions of data

An important component in the analysis of data is the identification of statistical distributions. To check how likely it is that an assumed theoretical distribution matches the actual historical distribution, statistical test procedures can be used to test hypotheses for their validity. Two hypotheses are formulated: the null hypothesis H_0 and the alternative hypothesis H_1. H_0 and H_1 are disjoint sets and together yield the set of total possible parameters (Holling & Gediga 2015, p. 25). In the statistical test problem, the following setting, which is also used in testing distributions, is called an *undirected problem,* where the above preconditions hold and θ stands for any parameter to serve as a test criterion for distinguishing the two sets:

$$H_0: \theta = \theta_0 \; against \; H_1: \theta \neq \theta_0 \tag{2.1}$$

Statistical tests contain a test statistic. This is a single value that is the basis for whether H_0 is retained or rejected. The null hypothesis H_0 is rejected at a given significance level if the test statistic exceeds a critical value. The significance level α describes the probability with which a null hypothesis is to be rejected even if it is true. Common values for significance levels are $\alpha = 0.1$; $\alpha = 0.05$ or also $\alpha = 0.01$ (Eckstein 2014, p. 311). The validity of H_0 is then tested for the (1- α) quantile.

The Anderson-Darling test (AD test) uses a different property of the normal distribution than the widely used Kolmogorov-Smirnov (KS) test and the X^2 test. While the latter are based on differences between the empirical and the theoretical normal distribution, the AD test is based on the analysis of the symmetry of the distribution. Since the normal distribution is axisymmetric around the mean μ, the extent to which this symmetry is observed can be tested for an empirical distribution. Razali & Yap (2011, p. 32) compare the AD test with the KS test and other hypothesis tests against different distributions and for different sample sizes. The KS test does not perform better than the AD test for any setting.

Given *n* observations sorted by their size $x_1 \leq x_2 \leq ... \leq x_n$, the test statistic of the test according to Anderson & Darling (1954) is then given by

$$AD_{emp} = -n - \frac{1}{n}\sum_{i=1}^{n}(2\mathrm{i}-1)(\log\left(F_0(x_i)\right) + \log\left(1 - F_0\left(x_{(n-i+1)}\right)\right) \tag{2.2}$$

when it is tested whether the observations are distributed according to the cumulative distributive function F_0. This test is initialized with

$$H_0: F(x) = F_0(x) \tag{2.3}$$

and

$$H_1: F(x) \neq F_0(x) \tag{2.4}$$

and can be performed with a comparison of the values for the test statistic AD_{emp} and the critical value for the test statistic AD_{krit} at a given significance level α. The AD test will be used in this thesis for the estimation of simulation input parameters.

2.3.4 The clustering algorithm DBSCAN

One data type that will be used in this thesis for identifying the structure and behavior of the production system is in-door localization data. To retrieve the necessary information about workplaces, walking paths, etc. from it, an efficient clustering algorithm is needed, which will be presented in the following.

Cluster analysis as a way to automatically sort data without prior knowledge is almost indispensable in today's world and is used in many areas, such as science, economics, medicine, and any other field where complex data sets have to be structured and classified. Cluster analysis deals with structuring a dataset A, with m elements having n ≥ 1 properties, to produce c groups, where $m, n, c \in N$. (Scitovski et al. 2021, p. 31)

Two important concepts for clustering are hierarchical and partitioning approaches. Hierarchical clustering algorithms can be divided into two subgroups, the divisive methods (top-down) and the agglomerative methods (bottom-up). In the agglomerative approach, first, each data point is considered as a separate cluster and then the two clusters with the smallest distance (which means greatest similarity) are merged. This step is repeated until a large cluster with all data points is created at the end. Divisive methods start with the whole data space as one cluster and divide it further and further until, theoretically, each point forms a cluster. Depending on when the merging or splitting of clusters is stopped, the desired number of clusters is obtained. (Scitovski et al. 2021, p. 81)

In partitioning methods, the data space is decomposed into non-overlapping, non-empty subsets, usually with a prior determination of how many k clusters the data set should be divided into. (Ester et al. 1996)

The DBSCAN, short for Density-Based Spatial Clustering of Applications with Noise, was developed by Ester et al. (1996) and is a bottom-up approach. The algorithm is density-based and can exclude noise points. Two of its strengths are its ability to detect arbitrarily shaped clusters and that it does not require a specification of the final cluster number. The DBSCAN algorithm has become increasingly popular over time and is now used in various fields, such as medical image analysis, spam detection, geography, etc. (Scitovski et al. 2021).

For an understanding of the algorithm, based on (Scitovski et al. 2021, pp. 93-97), some terms have to be defined first.

ϵ-Neighborhood of a point: Let $A \subset \mathbb{R}^n$ be an arbitrary data set of points and $\epsilon > 0$. Let the ϵ-neighborhood of a point $p \in A$ be defined as

$$\mathbb{N}_A(p, \epsilon) := \{q \in A \mid d(q, p) \leq \epsilon\} \qquad 2.5$$

where $d: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}_+$ is a distance function.

directly density-reachable: Let $\epsilon > 0$ and $minPts \in \mathbb{N}$ be predefined values. A point $p \in A$ is directly density-reachable from a point $q \in A$ if

$$p \in \mathbb{N}_A(q, \epsilon) \qquad 2.6$$

and

$$|\mathbb{N}_A(p, \epsilon)| \geq minPts \qquad 2.7$$

density-reachable: A point p is density-reachable from a point q if there exists a chain of points $p_1, \ldots, p_n$; $p_1 = q$, and $p_n = p$ such that p_{i+1} is directly density-reachable from p_i.

The basic idea of DBSCAN is that all points of a cluster contain at least minPts points in their ϵ -neighborhood, i.e., the density in the neighborhood must exceed a certain threshold. Therefore, $\epsilon > 0$ and $\mathrm{minPts} \in \mathbb{N}$ must be specified by the user beforehand. The algorithm assigns each point to one of the following three categories (Scitovski et al. 2021, p. 94):

A point $p \in A$ is a **core point** if its neighborhood contains at least minPts points of A, i.e., $|\mathbb{N}_A(\mathrm{p}, \epsilon)| \geq \mathrm{minPts}$.

A point $p \in A$ is an **edge point** if p is not a core point but is still density-accessible from a core point, i.e., $|\mathbb{N}_A(\mathrm{p}, \epsilon)| < \mathrm{minPts}$ | $\mathbb{N}_A$ and $\mathrm{p} \in \mathrm{B}\epsilon(\mathrm{p})$, where q is a core point. The set of all edge points of a cluster is also called the cluster edge.

Noise points are points that are neither core points nor edge points and therefore are not assigned to a cluster.

If two core points are directly density-accessible, they are combined into a cluster. Edge points are also added to a cluster if they are density-accessible from it. If edge points exist that are density-reachable from two or more clusters, they are added to any of the possible clusters. Usually, the algorithm is calibrated to add this edge point to the first discovered cluster. This makes the algorithm deterministic only if the data set is not permuted (Ester et al. 1996).

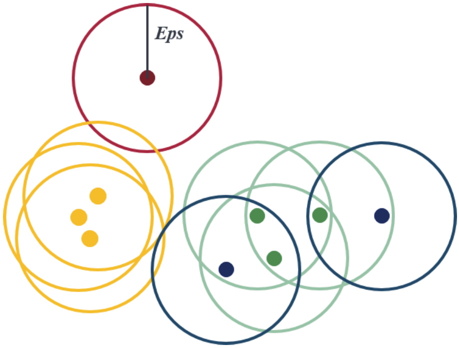


Figure 2-13 Visualization clustering with the DBSCAN algorithm

Figure 2-13 shows how a set of nine points is clustered by DBSCAN with parameter values $\epsilon = \mathrm{Eps}$ and $\mathrm{minPts} = 3$ into two clusters (yellow, green) and one noise point (red). Here, the circles represent the ϵ-neighborhood around each point. The green and yellow circles are the neighborhoods around the core points. These core points have at least three points in their ϵ-neighborhood. However, the points with the blue ϵ-neighborhoods are edge point, since their ϵ-neighborhood contain only 2 or 1 points, but they are still density-accessible from a core point. Therefore, they are assigned to the green cluster.

Three problems of DBSCAN are first, the specific input of $minPts$ and ϵ, second, the poor detection of meaningful clusters when the density of the data varies, and third, the high computational cost (Khan et al. 2014).

To sum up, the DBSCAN is an efficient clustering algorithm that offers several benefits for its application to recognize structures in localization data and was therefore chosen for this task in this thesis.

2.3.5 Process mining

One way to analyze enterprise data that takes a process perspective and shall be used to extract dynamic and structural information of the production system for its Digital Twin is process mining, which will be explained in the following.

2.3.5.1 Definition process mining

Process mining combines model-based approaches with data-centric approaches and aims to bridge the gap between data science (no consideration of the underlying processes) and process science (model-based without confirmation by data). The goal is to use event data to extract process information. In the context of process mining, it is also referred to as event data. (van der Aalst 2016, pp. 15, 24)

Table 2-3 Example Event Log

Case ID	Event ID	Timestamp	Activity
5346573	1	15-06-2009:09:32	Start Process B
5346573	2	15-06-2009:10:32	End Process B
5346574	3	15-06-2009:10:52	Start Process A

Process mining generally requires an event log, as shown in Table 2-3 as an example. Process mining methods can be divided into three categories based on their purpose. In **process discovery,** the objective is to construct a new process model from event data. In **conformance checking,** a model is compared to actual event logs to reveal discrepancies between modeled and logged behavior. In **process enhancement,** an existing process model is corrected or extended based on additional (normally more current) event logs. (van der Aalst 2016, p. 32)

2.3.5.2 Alpha algorithm

An important method for process model recognition is the alpha algorithm, which first derives the sequence relationships such as causality, direct succession, selection, or parallelism between the activities from the event log which includes multiple process instances (visualized in Figure 2-14). These relationships are defined between all activities and summarized in a relationship table also called the *footprint* of an event log. From this table, a Petri net is created as a process model in the next step. To achieve this, characteristic patterns have to be identified in the table, for example the identified *XOR-join* pattern in the figure on the bottom right. (van der Aalst 2016, pp. 167-177)

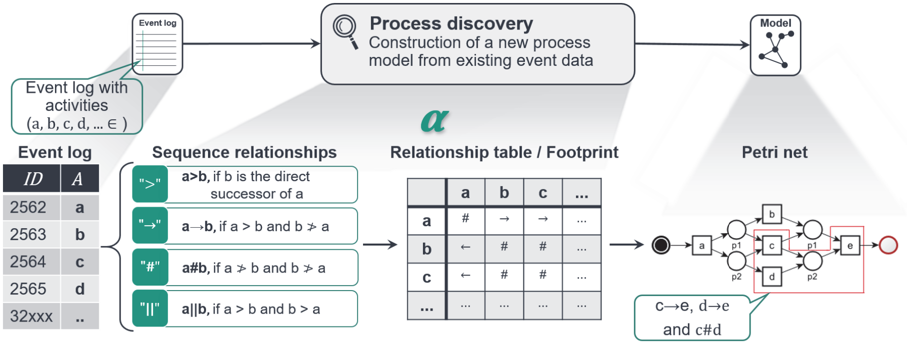


Figure 2-14 The alpha algorithm

Because of its easy applicability and its capabilities to recognize process information from common relational data in combination with filters, the alpha algorithm is used in this thesis to capture the material flow in the production system.

3 State of research

In the effort to develop a procedure for the development of Digital Twins of production systems, it is possible to refer to numerous preliminary works, which already cover some of the defined requirements from section 1.2. The relevant state of the art in research and technology is therefore summarized in this chapter and examined in terms of the requirements for a viable concept for Digital Twins of production systems arising from the motivation.

After an overview of publications on concepts for the creation and usage of Digital Twins, existing research on simulation input data management is presented. There are numerous preliminary works on (partially) automated or at least supported simulation model generation, as well as the consideration of real data in simulation models. In the following, the existing works are clustered according to the respective focus of the authors. The chapter closes with a comparison of the current state of research with the requirements, which follow from the motivation of this thesis in section 1.2. Not every cited work in this chapter can be compared to the posed requirements in a meaningful way, because some contain only classifications, theoretical concepts, or literature reviews.

3.1 Digital Twins in production research

The Digital Twin is currently a popular buzzword in science and industry, which is used in a variety of interpretations and contexts. Therefore, in the following, on the one hand, an attempt will be made to give the reader an insight into the wide range covered by Digital Twins through various classification approaches. On the other hand, relevant research work on the use of Digital Twins, especially in production systems, will be presented. All publications of this section (3.1) are either pure concepts or literature reviews and are therefore not suitable to satisfy the requirements from section 1.2 and a comparison would be useless.

3.1.1 Digital Twin classifications

In addition to the large number of definition attempts, there are numerous attempts to master the great diversity with which the term Digital Twin is used employing taxonomies and classifications, some of which will be presented here.

A broad taxonomy attempt for Digital Twins in general based on a literature review with 233 relevant publications is presented by van der Valk et al. (2020a). The results are 11 descriptive dimensions, but three of them remain unmentioned as they have the same expression for all publications. Thus, it remains a mystery what the unifying dimensions of Digital Twins are and eight differentiating dimensions remain with two to four possible expressions each: data link, purpose, conceptual elements between digital and real twin, accuracy, interface, synchronization, data input, and creation time. van der Valk et al. (2020b) develop another taxonomy of Digital Twins in simulative applications using the same approach. In this case, they classify Digital Twins according to the characteristics of the respective simulation component. In doing so, they use classical distinctions of simulations such as continuous vs. discrete, deterministic vs. stochastic, static vs. dynamic, terminating vs. non-terminating (see section 2.2.2) and thus sort 69 publications on Digital Twins from different fields.

Lechler et al. (2020) develop the structure model shown in Figure 3-1, which comprises three dimensions with their respective characteristics plus the application levels. To demonstrate the application of their structure model, they sort 20 published Digital Twins with it.

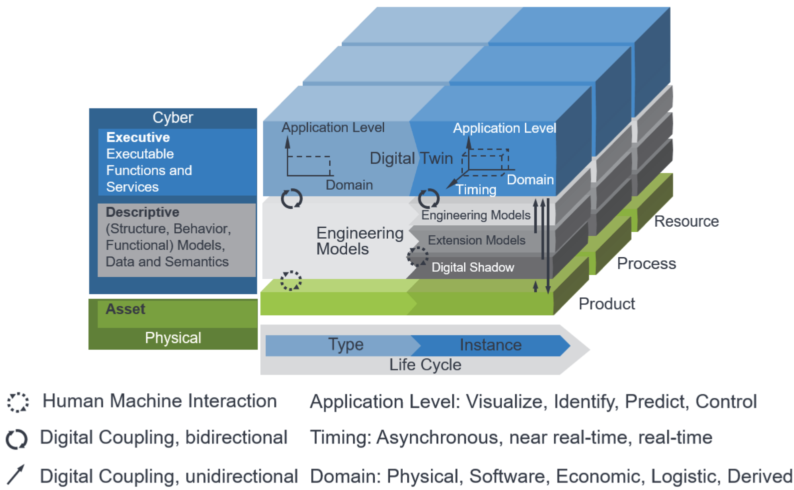


Figure 3-1 Structure model for Digital Twins according to Lechler et al. (2020)

Tao et al. (2017) extend the understanding of the Digital Twin from the three original building blocks (physical twin, virtual twin, and data link) to include the building blocks of data and services.

A literature review with an attempt to assess the relevance of individual publications on the topic of the Digital Twin, as well as the presentation of influential definitions and graphical representations of the concepts can be found in Sjarov et al. (2020). The publication also explains the related concepts of Digital Shadow, Digital Triplet, Product Avatar, Virtual Twin, and Virtual Twin Data Space by means of references to existing literature.

van der Valk et al. (2022b) identify in six interviews with experts from production and logistics eight requirements for data handling, data management, and services for the successful use of Digital Twins. In line with the goals pursued in this dissertation, these include the synchronization of the Digital Twin with reality, its simulation capability, its (semi-)automation, and the ability to share data via interfaces.

Again based on a structured literature review and 15 expert interviews, van der Valk et al. (2022a) define five archetypes of Digital Twins: Basic DT, Enriched DT, Autonomous Control DT, Enhanced Autonomous Control DT, Exhaustive DT, using the same methodology as before.

Kuehner et al. (2021) present a meta-review collecting and comparing 24 existing literature reviews on the topic of Digital Twins. In particular, they analyze the definitions and usage of related terms such as Digital Model and Digital Shadow, as well as the demonstrated benefits and as challenges of Digital Twins.

Another literature review of the concepts, technologies, and applications of Digital Twins is provided by Liu et al. (2021). In particular, they emphasize the simulation capability of Digital Twins and consider the potential applications in each phase of the system's life. They consider all types of Digital Twins.

Shao & Helu (2020) published a framework for Digital Twins in production whereby they address the various perspectives from the literature, which differ in the definition of the Digital Twin, the level of consideration (product, process, system), the targeted accuracy and the temporal integration (offline or real-time). They cite minimizing machine downtime, optimizing PPC, and virtual commissioning as possible benefits of Digital Twins. The authors announce the ISO Digital Twin Framework for Manufacturing standard (ISO/IEC 2023).

Fuller et al. (2020) collect enablers, challenges, and open research questions related to Digital Twins. Named use cases are smart cities, manufacturing, and healthcare while the open research questions include multidisciplinarity and standardization of Digital Twins, as well as challenges arising from inflated expectations and compatibility issues.

Bianconi et al. (2020) apply the design thinking approach to the existing Digital Twin definitions from literature and try to find their commonalities. These are the processing of historical data and sometimes data of sensors, coverage of multiple life phases and thus system levels, and simulation in the different life phases of real systems. The Digital Twin goals should not be part of the Digital Twin definition as they are evolving and highly use case dependent.

Another approach to structuring the field of Digital Twins, this time using 10 properties derived from a literature review, is provided by Autiosalo et al. (2020). The properties are data link, coupling, identifier, security, data storage, user interface, simulation model, analysis, artificial intelligence, and computation. This structuring is tested on seven application examples of Digital Twins from literature.

3.1.2 General concepts for the use of Digital Twins

Bao et al. (2018) propose a concept for the creation and operation of Digital Twins in production, attempting to cover product Digital Twins, process Digital Twins, and operation Digital Twins, as well as illuminating the interoperability between them. This is demonstrated using AutomationML on a machining center, resulting in significant cycle time reductions and more precise quality checks.

Riedelsheimer et al. (2020) present a ‘Digital Twin Readiness Assessment’ based on 26 expert interviews. The focus is on the further development of business models through the Digital Twin, the added value that can be generated and current concepts, measures, and capabilities for the Digital Twin. They observe that most current implementations in industry are not yet as mature as they are in research.

Kober et al. (2022) show based on literature research with 77 current application examples that currently no method for the definition of necessary degrees of accuracy for Digital Twins exists and propose their own methodology, without, however, demonstrating it with an example. It is embedded in an overall procedure to be able to make cost-benefit estimates for the creation of Digital Twins since higher accuracy usually leads to higher costs. For this purpose, independent elementary variables are derived from the target variables via intermediate variables, for which requirements are finally defined.

Redelinghuys et al. (2020) present a reference architecture for Digital Twins consisting of six layers. This is intended to meet the requirements for service-based, horizontal, and vertical integration in real-time, as well as for applicability, in both old and new production systems. The layers are (1) actuators and sensors of the physical twin, (2) controllers of the physical twin, (3) local data repositories, (4) IOT gateways, (5) cloud-based information repositories, and (6) emulation and simulation (overarching layers 3-5). SQL and OPC UA should be used for communication. This communication architecture is implemented for a robotic gripper.

A microservice architecture for Digital Twins with five components for virtualization, interoperability, data management, model management, and service management, including suggestions for open-source solutions to implement the components, is presented by Damjanovic-Behrendt & Behrendt (2019). They use an abstract example to show how this architecture could be implemented but focus on machine learning (ML) models rather than simulation models.

Uhlenkamp et al. (2019) attempt to systematize the use cases of Digital Twins based on a literature review. This results in the dimensions *goals, user focus, lifecycle focus, system focus, data sources, data integration level,* and *authenticity,* all of which have two to four possible manifestations.

Drivers, enablers, and barriers of Digital Twins in manufacturing are compiled based on six expert interviews and a literature review by Neto et al. (2020). External drivers are the need for flexibility in production, increasing competitive pressure, and the spread of the term ‘Digital Twin’ as a buzzword. Internal drivers are internal improvement projects, transparency goals, and employee safety. Enablers are needed in the areas of systems and technology, processes, people and competencies, and culture and strategy. Obstacles must be removed in all of these fields.

Klostermeier et al. (2019) investigate possible business models that can be driven by Digital Twins, such as ‘Equipment as a Service’ or ‘Predictive Maintenance’. In addition to the development of the Digital Twin, they state the formation of the customer's employees, i.e., the training of the subsequent users of the Digital Twin, to be an important future cost factor when it comes to Digital Twins.

Cimino et al. (2019) conduct an extensive literature review on use cases of Digital Twins in manufacturing with a particular focus on their integration with MES. They identify the key research gaps as the integration of the Digital Twin with the control system and the provided Digital Twin services because all implementations found to provide only selected services. They then introduce the Digital Twin of the learning factory at their institute. It can record and visualize energy consumption as well as PLC data, and track transport carriers. It has no simulation capability.

Tao & Zhang (2017) propose the Digital Twin shop floor, a concept for digitizing the shop floor that consists of four components: physical shop floor, virtual shop floor, shop floor service system, and shop floor Digital Twin data. They also describe its functioning before, during, and after production. Their focus lies on modeling and monitoring individual machines and products.

Magnanini & Tolio (2021) present their model-based System Digital Twin concept which is not based on DES but on a stochastic analytical model for performance evaluation which uses a set of partial differential equations. The corresponding hyperplanes are updated each time the production system configuration changes. Unfortunately, the authors do not provide insights into the validation and update mechanism. They demonstrate their concept at a real industrial use case and use it for buffer capacity evaluation.

3.2 Simulation input data management

Based on six use cases from two companies in the Swedish automotive industry, prepared with interviews, Bokrantz et al. (2018) describe quality issues with simulation data and the data generation process. From this, they derive numerous general suggestions for users to increase data quality for simulation projects. Barring et al. (2018) analyze challenges in obtaining data for simulation models using two examples and then call for common standards regarding data ownership, the relationship of data to key performance indicators (KPIs), and deriving information and decisions from data, for which they also propose five general guidelines. These two descriptive publications only provide general recommendations for simulation input data handling and are therefore not suitable for a comparison with the research requirements of this thesis.

Aufenanger et al. (2010) describe a concept for a generic interface for machine data integration into the simulation software d3FACT insight and its implementation. For this purpose, they use the Devices Profile for Web Services specification. Even though some parameters can be updated automatically, many requirements for a Digital Twin are not fulfilled.

Fritz (2007) develops in his dissertation a heuristic for the selection of suitable analysis methods to be used with simulation models based on a classification of the analysis problems in production planning according to necessary detail and investigation frequencies. Model structure, methods, tools, and generated output information are considered. In addition, an interface concept is developed and implemented using Microsoft Excel, which enables the semi-automatic simulation model construction from static models of the manufacturing system. He distinguishes between simulation-irrelevant, simulation-relevant, directly applicable, and indirectly applicable data. A systematic methodology for validation and update as well for the analysis of the resulting model accuracy is missing.

Müller-Sommer (2013) focuses in his dissertation primarily on the automated plausibility check of input data for supply simulations. He considers the quality, correctness, and completeness of input data to be critical factors for the economic generation of simulation models because the most costly tasks in model building are data procurement, plausibility check of data and modeling. By unifying different computer-aided methods in a simulation framework, the author creates a platform that cleans simulation-relevant data from production and makes them available in a simulation database for model building. This should reduce the effort for data plausibility checks and increase the efficiency of the simulation process. He does not consider validation of the model itself.

3.3 Validation and verification of simulation models in production

All cited publications on V&V do not include a real industrial use case and also do not discuss model correction or update in detail, which should be the logical consequence of a negative V&V result. Hence, a comparison with the posed requirements is not possible.

In a practical guide to support simulation modelers in simulation projects to ensure acceptance and credibility of simulation results, Balci (1989) describes 10 processes, 10 phases, and 13 credibility assessment levels. For this purpose, individual components such as model, data, and experiment design are first verified and validated, from which higher-level stages such as the credibility of simulation results and the acceptance of these follow. Based on this systematic, the author presented also an exemplary model V&V on a use case (Balci 1998).

Carson (1989) explains various V&V techniques in practice and gives an application example, while in 2002 the same author addresses several categories of modeling errors: project management errors, data and data model errors, logic errors, and experimentation errors (Carson 2002). Examples for these are given and metrics for V&V assessment are discussed.

Over three decades, Robert G. Sargent published his tutorial on simulation model V&V at the annual Winter Simulation Conference. In addition to presenting various validation techniques, he specifically addresses the different types of validity such as data validity, the validity of the conceptual model and of the implemented model, and operational validity. For testing the latter, examples of possible techniques he describes are the graphical comparison of the data, the comparison of the model behavior, calculation of confidence intervals, and hypothesis testing. It is supplemented by notes on validation documentation and a suggested validation procedure. (Sargent 1991)-(Sargent 2020)

In a similar tutorial in the same conference series, Robinson (1997) presented his view and findings on the V&V of simulation models. He distinguishes between white-box validation (consideration of the model interior) and black-box validation (consideration of only the model output).

Thaker et al. (2004) summarize considerations, definitions, concepts, and methods for general model V&V to develop highly accurate simulation models of nuclear weapons testing that can replace nuclear weapons testing.

Rabe et al. (2008, p. 131) explain that the assumption an automatic generation of the model leads to reduced V&V efforts is false and on the contrary, there may even be an increased V&V expense, as the automation of the generation process increases the risk to not detect errors in the data.

3.4 (Partially) automated simulation model generation and parameterization

In this section, first, some theoretical concepts and literature reviews concerning (partially) automated simulation model generation and parameterization will be presented and discussed that are not suitable for a comparison with the research requirements.

A literature review of recent work on automated simulation model generation (ASMG) with a focus on information retrieval was conducted by Reinhardt et al. (2019). They note that most existing approaches still require data formalization. Recent works seem to focus more on information retrieval from dynamic data using, for example, location data or ML. The use of complementary sensor data in addition to existing IT systems is also increasing. This is a trend that will also be addressed in the work presented here.

Scheer et al. (2021) present and compare different possibilities to couple simulation models and reality. For this purpose, they consider the concepts of manufacturing data analysis, offline simulation, online simulation, dynamic data-driven application systems, symbiotic simulation, Digital Twins, and cyber-physical systems (CPS). Their distinguishing features are the connection direction between the real and virtual world, as well as the usage of simulation and the existence of an integrated computing unit.

Rabe et al. (2008, p. 131) distinguish between three categories of ASMG:

1. generation of layout data for the simulation model based on an existing CAD layout
2. generation of models with the help of work plan data
3. generation of executable based on existing (non-executable) models or model descriptions

As mentioned before, the objective of this thesis is not to automatically generate models, but to keep existing models accurate. Anyway, the three categories of ASMG from 2008 are not sufficient to describe modern ASMG approaches since they rely on much more and diverse data.

Further, purely theoretical contributions and concepts on automatic data input to simulation models or their automated generation include Kotiades (2016) and Onggo et al. (2020).

Besides these purely theoretic publications, various publications exist which include the description of methods, technologies, and concepts for automated model generation and parametrization that could be directly applied to a real use case. These works, which will be introduced in the following, partly or fully satisfy some of the requirements on a Digital Twin concept for production systems presented in section 1.2. Therefore, they will be included in the comparison table at the end of this chapter.

Rooks (2009) presents an automatic generation of logistics simulation models from planning data. The author shows that the existing database is not sufficient to create a valid model, which is why the data in the planning systems had to be enriched for this purpose. The author does not describe any validation or update after the model generation.

In order to couple a simulation model for simulation-based real-time control with real-world data, Horn et al. (2005) develop two strategies to pass the actual state of production to the model, and study the performance of both. Information on dynamic behavior and structure of the system are neglected.

Santos et al. (2022) investigate the use of DES as a Digital Twin of a non-automated intralogistics process within a plant. A simulation model in the software FlexSim is regularly updated with the material consumption at each kanban supermarket. The authors refer to this as near real-time. The simulation model is then used by logistics planners to define routes, for which a graphical user interface for the simulation model is developed. Only little information is updateable and update is repeated in fixed intervals without prior validation.

Heitmann (1999) proposes stochastic compensation models for higher fidelity of simulation models to reality. The compensation models result from the evaluation of the coherent behavior between real and simulated production behavior using statistical methods such as regression and variance analysis. While real data is used for validation, the approach does not include any update with real data.

Mieth et al. (2019) use in-door localization data as input in a material flow simulation and present a framework of how to rely a simulation model entirely on this data source, to have a single source of truth. Therefore, they check the localization data for data quality dimensions and enumerate which information, necessary for material flow simulation, can be theoretically gathered from localization systems. While a lot of information can be obtained this way, the authors do not discuss the necessary mechanisms for validation and update or the consequences on the simulation models accuracy.

An older contribution to automatic model generation from data exemplarily carried out at a model factory, can be found in the dissertation of Eckardt (2001). Since he is explicitly focusing model generation, validation, updates, and the resulting Digital Twin behavior are not discussed.

Inspired by various forms of symbiosis in biology, Aydt et al. (2008) present five different forms of symbiotic simulation systems: as a decision support system, a control system, a prediction system, a model validation system, or for anomaly detection. They then present a hybrid symbiotic simulation using an example from semiconductor manufacturing which is only partly implemented and only with an emulator instead of the real system. Input computation, validation and update are only discussed roughly.

An XML-based approach to input design data (specifically the Piping and Instrumentation Diagram) as a contribution to the ASMG of plants in the process industry using the meta-data exchange format CAEX (computer aided engineering exchange) was proposed by Barth et al. (2009). With this approach, planned parameter values and structure information can be automatically integrated into the model. Updates shall be performed in case of a change in planning. A further examination of the resulting model accuracy and its dependency on data quality is missing.

A data-based approach for ASMG of assembly line from CAD layout data with the associated logistics was presented by Wy et al. (2011). Among other things, they use the Bill-of-Material as well as other planning data. Like many other ASMG approaches, they do not discuss a proper concept for model validation and update and do not analyze the resulting behavior of the simulation model.

In the final report of the research project 'DigitTwin', after some conceptual work and exemplary use cases of data collection, analysis and modeling, Stjepandic et al. (2022, p. 243) provide a research outlook that focuses primarily on the necessary updates of the Digital Twin in the event of changes in the production system, which was apparently not considered in the project. They distinguish between parameter updates, structural adjustments, and the complete restructuring of the Digital Twin. While their report touches many concepts and ideas, they do not provide one complete applicable use case and do not perform a further analysis of the resulting Digital Twin.

As early as 2002, Werner and Weigert (2002) published an approach for the optimization of production planning based on process accompanying simulation models. For modeling, the authors use ROSI, an object-oriented tool for discrete-event simulation. A model template which was created in particular for electronics production is fed by production-relevant data from ERP and Production Data Acquisition (PDA) systems. To map the relationship between simulation and production planning, a theoretical time model was formulated that divides production into two process levels (planning and production). Planning is modeled by a virtual time component and production by a real-time component. Finally, the application of an error metric (error in terms of duration and start time of an event) allows continuous analysis and monitoring of the discrepancy between the simulation model and real production. In case of a discrepancy, the existing rather abstract model can thus be quickly synchronized and adapted to the real status of the production system during the process. Their description of the use case, methods, and results remains short. This makes their approach unreproducible even though the graphs they show look promising.

Kirchhof (2016) briefly describes the use of ERP and MES data to automatically create a simulation model using a template for production lines in the SIMIO software. The data are collected from the systems using custom-build software. Due to the compactness of the paper, many aspects remain open. How his approach satisfies the posed requirements has to be inferred from the provided information.

Jensen (2007) proposes in his dissertation a methodology for the partially automated generation of simulation models for material flow systems. Jensen uses the data format XML, in a first step, to transfer heterogeneous data from different IT systems of the production environment into a uniform format. The unified model data forms the basis for the simulation and can be integrated into the model-building process in a partially automated way. This is intended to provide the user with a generic base model, which can be expanded in an intermediate step into a rough planning model and ultimately developed into more complex models. The methodology is intended to ensure closer integration of material flow simulation with production planning and, at the same time, to reduce the effort required for model generation. While this approach includes the recognition of several aspects of the simulation model information base from planning data for automated generation, many aspects can still not be recognized with this approach. Mechanisms for model validation and update as well as an in-depth analysis of the resulting model are missing.

Goodall et al. (2019) develop a simulation model with automatic data input to support remanufacturing operations based on a generic core model. Among others, data from radio-frequency identification (RFID) scanners are used. These provide data for the estimation of parameter but not about its structure or dynamic behavior. Even though the motivation of the authors is to reduce the costs of maintenance and update of simulation model and therefore to advance the topic of Digital Twins, they do not discuss the validation, update or behavior of their model and focus solely its input data connection as well as exemplary experiments.

Ding et al. (2019) and Vachalek et al. (2017) exemplify numerous studies that construct Digital Twins in small, controlled experimental environments with limited data input. Ding et al. (2019) use the concept of cyper-physical production systems on production resource level to obtain, process, and save data of each resource with focus on parameter and structure information. The authors do not perform an analysis of the achievable accuracy of the Digital Twin. Vachalek et al. (2017) use an OPC data server to input parameter values from a small automatic assembly station with manual part loading to a Plant Simulation model. Input computation, model validation and update, and system examination are not discussed.

3.5 Core Manufacturing Simulation Data (CMSD)

The Simulation Interoperability Standards Organization (SISO) developed the Core Manufacturing Simulation Data (CMSD) standard to standardize data exchange between enterprise and manufacturing IT systems and simulation software (SISO 2010). This was used for some scientific work in the following years but is not used in industry and research today. For example, Skoogh et al. (2012) presented a system for generic data management that is capable of outputting arbitrary data from various sources via the steps of data extraction, data transformation, and output preparation in the Extensible Markup Language (XML) format according to the CMSD standard. Their focus lies primarily in the preparation of model parameters while other aspects of simulation input information are neglected. While they do discuss the model update, the necessary validation is not subject of their research.

In his dissertation, Barlas (2015) developed open-source software for automatic data input into DES, where data from enterprise IT systems are translated into CMSD format for this purpose. He does not exclusively discusses parameter estimation but also the dynamic behavior and structure of the production system. His research ignores model validation and an analysis of the resulting simulation model.

Bergmann (2013) used in his dissertation the CMSD standard for automatic model generation after a comparison with alternative standards like SDX (Simulation Data Exchange), STEP (Standard for The Exchange of Product model data), SDL (Specification and Description Language), and SysML (Systems Modeling Language). This is done in successive use cases of a workshop production. In the developed overall architecture, simulator-specific model generators are required, of which two are created for the Plant Simulation and SLX simulators. Interfaces to enterprise software are discussed only superficially. Validation was performed in both laboratory and field experiments. His approach is quite exhaustive on the data input side but fails to cover sensible mechanisms for model validation and update after model generation. He shows how parameter values can be integrated into the model, but not how data preparation and parameter estimation can be performed.

Fournier (2011) presents the CMSD application programming interface (API) and various translators that used it. To prove their functionality, a CMSD file is translated into different simulation programs. His publication focuses on the technical transfer of the data into the simulation programs and does not discuss input computation or the logic how to design model validation or update. CMSD-based approaches have not been found in practice as the CMSD standard has not yet become widely accepted (Bergmann & Straßburger 2020).

3.6 Detection of dynamic behavior

Selke (2005) focuses in his dissertation on the automated integration of processes and strategies of PPC into the generation of operations accompanying simulation models. The aim of his dissertation is to support short-term decision processes in a dynamic production environment with the help of such a simulation model at low cost by using it as an adaptive planning and forecasting system during operation, for example for order release or parameterization of flow rules. For this purpose, the author first combines different approaches for the description and identification of strategies and processes of production control into a metamodel. This metamodel serves as a basis for the automated interpretation of sequences and strategies using the method of pattern recognition in operational data. Finally, this metamodel is integrated into an automated procedure for simulation model building. The implementation is done with Microsoft Excel and Microsoft Access with data from PDA and the production control system. The focus of his research is the integration of dynamic behavior into the model, while parameters and structure are neglected, as well as mechanisms for model validation and update. No in-depth analysis of the accuracy of the simulation model was performed.

Like Jensen (2007), Zenner (2006) uses XML to enable automatic model generation. In this case, the material flow simulation is primarily used for variant planning. He describes that the automatic model generation from process graphs, which are supplemented by resource information, is only possible under very narrow boundary conditions, which is mainly due to the insufficient representation of process logic. He tries to solve this deficit with two generic logistics modules and one variant module. In this way, material flow control and parallel processing steps can be mapped. For the implementation, he also wants to rely on commercial simulation software solutions and therefore performs a comparison of two solutions. He proposes heuristics to ensure that station modules are positioned in the layout in a way that they do not overlap and that material flows cross each other as little as possible. The author’s research touches multiple aspects of simulation input computation but does not elaborate on model validation and update or the achievable accuracy of the simulation model.

Nagahara et al. (2020) describe how the parameterization of abstract agents, which are responsible for determining process times and order release in a simplified simulation model of a production line, can be done automatically with the help of production data if these are not granular enough to recognize everything necessary from them directly. For this purpose, the unknown parameter and decisions are modeled by ML agents who are trained to act as realistically as possible. This is done for two simple examples (with one and two machines, respectively). This innovative approach is briefly introduced and not deeply analyzed so that many relevant aspects for a complete Digital Twin approach remain open.

Rozinat et al. (2009) take a purely data-driven approach and combine various aspects of process mining in their work to automatically create a simulation model as a Colored Petri Net based on event logs, which is thus rather abstract and poorly visualized compared to commercial simulation software. The simulation model includes available resources, capacity constraints, and control loops. The developed approach is validated with the help of the process mining framework *ProM 4.0* and two case studies.

A theoretical consideration of how simulation and process mining can be used together is provided by van der Aalst (2018). Due to its purely theoretical nature, it will not be included in the research requirements evaluation at the end of this chapter.

Bergmann et al. (2017) summarize their previous activities for the automatic reproduction of dynamic processes or rules. They distinguish between manual mapping, matching of real data with a set of previously defined rules (e.g., by pattern recognition), and learning methods that try to imitate the systems’ behavior. The authors see the latter as particularly promising. Thus, Bergmann et al. (2015) investigate how dispatching rules could be detected using data mining methods. In particular, they investigated Naive Bayes Classifier and Support Vector Machines for this purpose. Approaches to inferring behavior using artificial neural networks are already presented by Bergmann (2013). A comparison of the methods Decision Tree, Naive Bayes Classifier, Artificial Neural Network, Support Vector Machine, and k-Nearest Neighbors for production sequence inference of a production line based on a simple example is presented by Bergmann et al. (2017). Both publications focus strongly on the description of rules and dynamic behavior of the production system where they do provide new insights, but do not discuss parameter or structure recognition. Mechanisms for model validation and update are also neglected.

3.7 Detection of system structure

Splanemann (1995) deals in his dissertation with the recognition of the layout of the production from CAD data in STEP format. What is striking is his focus on the representation of the layout, which is now much easier to achieve with current software nowadays than at the time his work was written. Paprotny et al. (1999) also use a CAD layout file to automatically input the physical system components into a simulation model of an automated material handling system. Both approaches try to automatically capture the structure of the production system with consideration of its parameter or behavior. Because their objective is automated model generation, they do not describe a repeated model validation and update process.

Taking a comprehensive approach with respect to different levels of detail, Thiers et al. (2016) automate simulation model generation in commercial software for an aircraft manufacturer by enabling translation between different models with different levels of abstraction. Thus, there is no longer a distinction between parameter and structural changes. However, modeling is still necessary - at a higher level of abstraction. This places even higher demands on the modeler and user than the classical modeling in modern simulation software. Their approach covers the integration of parameter values (without data preparation and computation methods), structure and behavior information. A procedure to keep the model up-to-date is not presented.

Even more abstract is the model developed by Terkaj & Urgo (2015). The ‘Virtual Factory Data Model’, which can be used to generate abstract simulation models of production systems, uses an ontology and real data maintained according to this ontology. The authors extend their model to describe the relationships of different levels of granularity in simulation models using the mathematical method of delta lenses in Terkaj et al.

(2021). For this purpose, production sections consisting of several machines and buffers are combined into black boxes, their behavior is approximated and the influence of abstraction on the simulation accuracy is evaluated. This approach allows the automated integration of a lot of input information if the data is provided in the predefined format. Since the authors focus model generation, they do not describe mechanisms for model validation and update and do not perform any analysis of Digital Twin behavior.

3.8 Research gap

As shown above, the integration of real data is a frequently discussed and studied topic in the disciplines dealing with the use and further development of simulation software. Most approaches are strongly use case oriented and only consider a certain part of the data needed for simulation. In addition, there is usually insufficient validation and often using only artificial data. Additionally, the focus is mostly on the completely automated generation of simulation models and not on their maintenance.

Table 3-1 shows a comparison of the publications, which represent the state of research and were introduced earlier in this chapter with the requirements which arose from the motivation given in chapter 1. These publications were selected based on an extensive literature review and represent the most important works on this issue. They illustrate the various subjects which have to be treated when tackling the issue of Digital Twins of production systems as well as the manifoldness of directions from which this objective can be approached. For each work, the extent, to which each requirement is considered, is depicted using circles. If the requirement is fully fulfilled, the circle is completely black. A circle half-black, half-white represents that this requirement is discussed in the publication but not completely resolved. A white circle stands for the neglect of a requirement. Intermediate stages are expressed by circles where one or three quarters are filled. Are more concise version of Table 3-1 was published in (Overbeck et al. 2023).

Table 3-1 State of research

Section	Requirement	1. data acquisition and processing			2. procedure for validation and adaptation			3. examination of Digital Twin		
Section	**Aspects of requirement / Approaches as presented in chapter 3**	1.1 Parameter	1.2 Dynamic behavior	1.3 Structure	2.1 Automated validation	2.2 Automated update	2.3 Directly applicable to real data	3.1. Validation on real use case	3.2 Accuracy over time	3.3. Influence of available data
Data Input	Aufenanger et al., 2010	◑	○	○	○	◑	◔	○	○	○
Data Input	Fritz, 2007	◑	◑	◔	○	○	○	○	○	○
Data Input	Müller-Sommer, 2013	◕	◑	◑	○	◑	◕	●	◑	○
Model generation and parametrization	Rooks, 2009	◕	○	◑	○	○	○	●	○	○
Model generation and parametrization	Horn et al., 2005	●	○	◑	◑	◕	◕	●	◑	◔
Model generation and parametrization	Santos et al., 2022	◑	○	○	○	◑	◑	●	○	○
Model generation and parametrization	Heitmann, 1999	○	○	○	◑	◑	◑	○	◑	◕
Model generation and parametrization	Mieth et al., 2019	◑	◑	◑	○	○	◑	○	○	○
Model generation and parametrization	Eckardt, 2001	◑	◔	◑	○	○	○	○	○	○
Model generation and parametrization	Aydt et al., 2008	◑	○	○	○	○	◕	○	○	○
Model generation and parametrization	Barth et al., 2009	◑	○	◕	○	◑	○	○	○	○
Model generation and parametrization	Wy et al., 2011	◕	◑	◑	○	○	○	◑	○	○
Model generation and parametrization	Stjepandic et al. 2022	◑	○	◑	○	○	◑	◔	○	○
Model generation and parametrization	Werner & Weigert, 2002	◕	○	◑	◑	◑	◑	●	◑	◑
Model generation and parametrization	Kirchhof, 2016	◑	○	◔	○	○	○	●	○	○
Model generation and parametrization	Jensen, 2007	◑	◑	◔	○	◔	◑	●	○	○
Model generation and parametrization	Goodall et al., 2019	◕	○	○	○	○	◑	◑	○	○
Model generation and parametrization	Ding et al., 2019	◕	◑	◑	◔	◑	◑	○	○	◔
Model generation and parametrization	Vachalek et al., 2017	◑	○	○	○	○	◕	◑	○	◑
CMSD	Skoogh et al., 2012	◕	○	○	○	◕	◕	●	○	○
CMSD	Barlas, 2015	●	◔	◔	○	◑	●	◑	○	○
CMSD	Bergmann, 2013	◕	◔	◑	○	◔	◔	◑	○	○
CMSD	Fournier, 2011	◑	◔	◔	○	○	◔	○	○	○
Dynamic behavior	Selke, 2005	○	◑	○	○	○	◑	●	○	○
Dynamic behavior	Zenner, 2006	◑	◑	◑	○	○	○	●	○	○
Dynamic behavior	Nagahara et al., 2020	○	◑	○	◑	◑	◔	○	◔	◔
Dynamic behavior	Rozinat et al., 2009	◔	◑	◑	○	○	◕	●	◔	◔
Dynamic behavior	Bergmann et al., 2015	○	◑	○	○	◔	◔	◔	○	○
Dynamic behavior	Bergmann et al., 2017	○	◕	○	○	◔	◔	●	○	○
Structure	Splanemann, 1995	◔	◔	◑	○	○	○	◑	○	○
Structure	Paprotny et al., 1999	○	○	◑	○	○	◑	●	○	○
Structure	Thiers et al., 2016	◕	◔	◕	○	○	◑	●	○	○
Structure	Terkaj et al. 2015 + 2021	◑	◔	◑	○	○	○	○	○	○

● = completely fulfilled ◑ = discussed ○ = not considered

When looking at Table 3-1, it becomes obvious that the current state of research presented above does not adequately answer the necessary research requirements defined in chapter 1. Therefore, in order to enable the Digital Twin of the production system to be used on a broad scale, further research efforts are necessary, to which the present work is intended to contribute.

For the creation of a holistic concept for Digital Twins of production systems, the following open research questions need to be answered:

RQ1 Which data are needed and how must they be prepared?

RQ2 Which procedure for validation and update leads to an accuracy improvement and automated maintenance of the Digital Twin?

RQ3 How does a Digital Twin of a production system respond to changes in reality and data quality?

4 Own approach

The approach developed in this thesis aims to fill the previously identified research gap and to answer the identified research questions. In doing so, the scope of the approach is restricted to discrete-event material flow simulations of production systems, as defined in section 2.2.6. Many of the steps, ideas, components, and even presented results will be (partially) transferable to other types of simulations, where they may support the development of realistic Digital Twins but this is not the pretension of the present work.

4.1 Overall concept

To realize the goal of automated maintenance of the simulation model, the initial model, the so-called ‘seed model’, and the general Digital Twin procedure have to be defined first, before the individual components are elaborated in more detail afterwards. The method is based on the research of the author. Supporting work was done in student theses that were written under the guidance of the author and are cited with the scheme (A_<last name>, <year>). The overall approach was first described in general terms by Overbeck et al. (2020).
The general functionality of the developed Digital Twin of the production system is shown in Figure 4-1, which also indicates the two main steps of the approach ‘model validation’ and ‘model update’, which will be presented in detail in the further course of this chapter.

4.1.1 Digital Twin procedure

Starting from a seed model (see section 4.1.2), which is initially created manually by a simulation expert, the cycle of validation and updating begins, which is intended to maintain and possibly even improve the accuracy of the Digital Twin over its life cycle. For this purpose, the simulation model is first compared with the real system (validation). If the resemblance to reality is too low (negative validation), an update of the model is triggered. The first step of the update is to determine which component of the simulation model needs to be adjusted. The next step is to check how the necessary adaptation can be carried out. The adaptation can be carried out anyhow between fully automatically without any user intervention or with the guided involvement of the user. After the update, the validation is repeated. If a further adjustment is necessary because the validation is still negative, another component of the simulation model is updated in the next update step.

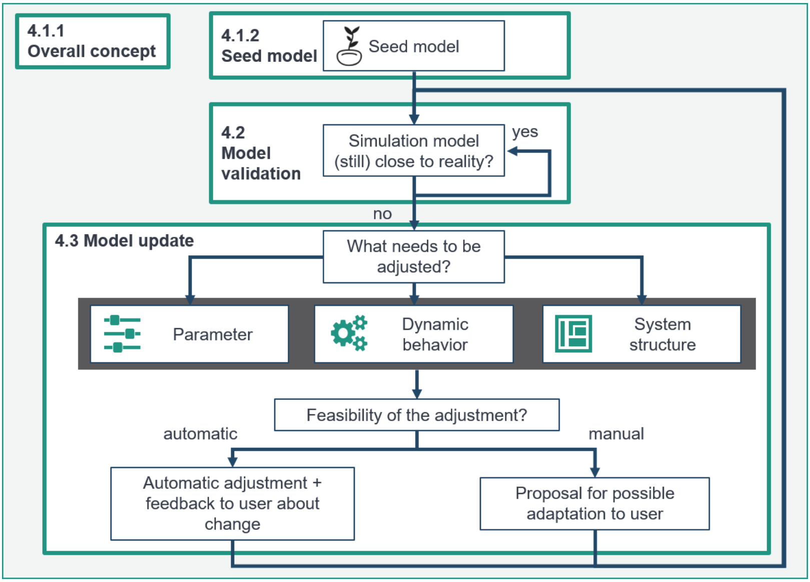


Figure 4-1 Function sketch of the Digital Twin of the production system including the respective text sections

In the presented Digital Twin procedure, real data from production are needed at four different moments:

1. parametrizing the seed model (section 4.1.2.2)
2. configuring the validation simulation runs (section 4.2.5)
3. comparing the real system behavior with the behavior of the simulation model after completion of the validation simulation runs to calculate accuracy metrics (sect. 4.2.6)
4. determining the new input information for the simulation model during the update (section 4.3.4).

Since the Digital Twin is more than a conventional simulation model, as described above, but in fact such a model constitutes the core of the Digital Twin, the terms ‘model’, ‘simulation’ and ‘simulation model’ are used in the following to refer to the underlying, initially created and continuously adapted simulation model in a suitable simulation software package. The term ‘Digital Twin’ refers to the entirety of the simulation model, data management, data preparation, validation, and updating mechanisms.

4.1.2 Seed model

The manually created seed model is the starting point from which the Digital Twin is developed using the methods described in this work. This approach was chosen because the previous approaches to automatic simulation model generation were not successful or required upstream manual modeling in some different form (see section 3.4). Furthermore, the construction of a rough, prototypical model of a production system with commercially available simulation software is significantly less time-consuming than the preparation of input data and the subsequent iterative improvement of model quality. Surveys have shown that especially the steps of data collection and preparation, as well as the ‘fine-tuning’ of the model are much more time-consuming than the actual modeling (Acél 1992; Müller-Sommer 2010). Besides, because modeling in particular requires human creativity and abstraction skills, this step can only be automated to a limited extent. The automation of other steps of simulation model creation and maintenance (especially the data-based ones) appears to be more promising for automation.

When creating the seed model, two steps must be taken into account, which in theory follow one another (e.g., (VDI 2014)), but in practice usually blur together: modeling and implementation. In both steps, essential aspects for a successful implementation of the Digital Twin have to be considered.

In the best case, a model is already created in the planning phase of the production system. If this is the case, it should be designed in such a way that it can be used as a seed model and transferred to be a Digital Twin of the production system during construction, commissioning, and ramp-up of the production system. This can be referred to as ‘greenfield approach’ and would be the perfect implementation of the vision shown in Figure 1-1. In the two use cases presented in this thesis, the Digital Twin was both times created for an already existing production system, which can be called ‘brownfield approach’.

4.1.2.1 Seed model modeling

Figure 4-2 shows some of the most important choices to make during seed model creation. In modeling, the first step is to understand the use case, e.g. the production system at hand. The next step is to select the appropriate level of detail and the modeling approach. The system boundaries must also be chosen wisely.

As in common simulation modeling, it is important to involve all stakeholders at an early stage to clarify long-term requirements and to get to know the different perspectives on the production system. It is of greater importance than in normal simulation model creation to already consider the long-term development and scenarios of the production system until the end of its life, dismantling or further use of the production system and to take these into account while modeling. This requires discussions with more strategic departments of the company (business analysts, network and supply chain planer, or strategy departments) that are normally not included in common simulation projects. A widespread way to capture the possible long-term evolution of the production system would be through scenarios. The model has to be designed to be adapted to the most probable or otherwise most important scenarios.

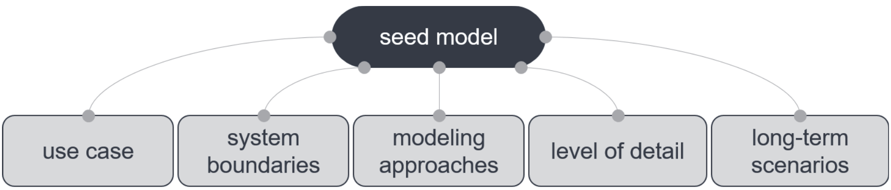


Figure 4-2 Choices to make during seed model creation

To lengthen the possible usage, it is advisable to focus on the flexibility of the modeling so that later adaptations (the need for which will be revealed by the validation mechanism of the Digital Twin) can be easily implemented. The use of generic building blocks (Brützel et al. 2020) or the recourse to a standardized library for the simulation software used (A_Maul 2021) support the flexibility of the model. The demand for flexibility can lead to additional effort in model creation and implementation, but this is justified, if it leads to longer and broader usability, from which an overall greater benefit follows that will outweigh the extra effort.

4.1.2.2 Seed model implementation

After the appropriate design of the model, some aspects have to be taken into account when implementing the extensible seed model. In order to successfully implement the described approach for Digital Twins, the model database must be designed in a modular way so that it can be easily and selectively automatically adapted in the following model maintenance process. Furthermore, the representation of the processes and decisions in the model must be designed in a way that they can be easily adapted and automated as far as possible. This can be achieved, for example, by avoiding fixed programming of the internal processes and reverting to tables in which the process steps are stored.

The presented approach assumes that the seed model is created in a commercial 'off-the-shelf' simulation software for discrete-event simulation of material flows in production and logistics. 19 examples of such software solutions were compiled by Dias et al. (2016). The advantages of using commercial simulation software to create Digital Twins include that internal mechanisms are already extensively tested and robust as well as, that visualization, which greatly simplifies model creation, validation, and evaluation, is provided. Furthermore, most commercial simulation software have broad capabilities for modeling a wide variety of conditions. Their widespread use leads to greater availability of experts (internal or external to the company) familiar with the software, as well as mature documentation for familiarization with it, even for beginners, and further assistance in using the software. (Law 2015, p. 182)

Next, the model must be initially parameterized, whereby the initial parameter set must be selected as close as possible to reality to accelerate the process of improving accuracy. This can be done in cooperation with experts familiar with the production system. Optimally the initial parameter set is created in the course of the initial development of the methods for simulation input calculation needed for the automated update (see section 4.3.4). If possible, real data should be used.

After creating the seed model, the next step is to develop the necessary mechanisms for approximating the simulation model to reality. These are core functionalities of the Digital Twin. As described in section 4.1.1, the procedure can be divided into two steps: the validation and the update. The steps are already rudimentarily described by Overbeck et al. (2021c).

4.2 Model validation

To make statements about the current fidelity of the simulation model to reality and to be able to derive decisions from it, as well as to monitor the success of the model maintenance process, procedures for the automatic comparison of the model with the real system are required. From the wide range of validation and verification (V&V) techniques described in Gutenschwager et al. (2017, p. 208) and Rabe et al. (2008), the V&V technique 'comparison with recorded data' is used for the developed approach due to its great informative value.

4.2.1 Validation process

As shown in Figure 4-3, the validation process consists of six sequential steps. Some settings for the validation must be made in advance. The process is first explained in general terms and individual aspects are discussed in greater detail in the following sections.

First, the validation period (in the past) has to be defined, taking into account various parameters such as data availability and quality, inherent system variations, but also the required computing time and power for data processing. Additionally, the number of simulation experiments to be performed (simulation runs with different starting values for the calculation of random numbers) has to be defined for each validation period.

Once the basic settings have been made, the necessary configuration data for the validation simulation runs is retrieved from the company databases (step 1). The data required for validation can be divided into two categories: firstly, information required for setting up the simulation for the validation runs, such as production plans, planned shutdowns, available employees in the period under consideration, etc. (referred to in the following as *validation input*); secondly, information on the performance and behavior of the real system in the validation period (referred to in the following as *reality output*). Part of the validation input are also exceptional events that occurred in reality during the validation period but are so rare that they are not explicitly represented in the model. Yet they have a decisive influence on the behavior of the system and would therefore prevent successful validation if they were not taken into account (see section 4.2.5.2) (step 2). Subsequently, the validation simulation runs are performed by the simulation software (step 3). If possible, the individual runs should be performed by the simulation software in parallel on different processor cores to speed up the validation process.

After the simulation runs, the corresponding results (*reality output*, e.g., number of produced parts) of the real system are queried (step 4) and the accuracy metrics are calculated to quantify the deviation of the simulation runs from reality (step 5). If the accuracy metrics violate the defined thresholds, the validation is considered negative and the update is triggered. Otherwise, the model is accurate enough and can be used for PPC as well as improvement work (step 6).

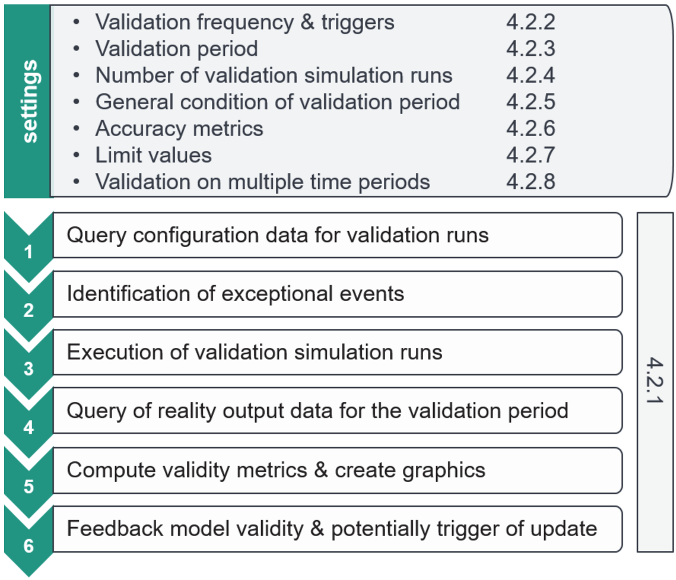


Figure 4-3 Validation process and previous settings

4.2.2 Validation frequency and triggers

The Digital Twin is supposed to represent an up-to-date image of the real production system when it is used, so validation immediately before its use for production planning or improvement work is obvious. However, because the validation and update cycle, which may have to be run through several times, can take some time, this would lead to a delay in the readiness of the Digital Twin and thus to a restriction of its usability. To avoid this, a periodic, usage-independent triggering of the validation is advisable. For example, a weekly or monthly routine validation of the Digital Twin would be conceivable.

4.2.3 Validation period

The determination of the considered validation period depends on the inherent fluctuation of the production system, the temporal structuring with which the production system is controlled (e.g. processing of weekly production plans, shift agendas, daily quotas), as well as the period length over which prognoses shall be made in the future. The length of the validation period should in any case be in a reasonable relation to the cycle time of production and the average length of failures to cover a sufficient amount of events. To fulfill these requirements one week should normally be the lower limit of reasonable validation period length. Longer validation periods can always be disaggregated into shorter time windows during the calculation and analysis of the accuracy metrics (see section 4.2.6), for example to gain insights into the validity for shorter periods or to quantify accuracy fluctuation.

4.2.4 Number of validation simulation runs

To correctly capture the range of variation of the simulation model (due to the implementation of stochastic processes and probability distributions), several simulation runs with different start values for the pseudo-random number calculation must be executed per validation period and subsequently compared with each other both in their final results and their progressions. It is not possible to give a general answer on the required number of simulation runs in particular, but it is recommended to start with a large number in the first validation runs in order to capture as many possible outcomes as possible. If feasible, the number should be reduced in the further course of use after several validation and update cycles, since the validation simulation runs can be time-consuming and can represent the most expensive part of the validation and update process in terms of time and computing costs.

There is also a correlation between the length of the validation period considered and the number of simulation runs required. Since a longer validation period already represents more contingencies, but also requires a longer computing time, fewer but longer runs can be sufficient.

Another aspect to consider is the number of simulation runs that are performed when using the Digital Twin to evaluate scenarios or actions. For example, if the internal standard for simulation experiments is at least 10 repetitions, the number of validation simulation runs should be based on this.

One possibility to determine the number of necessary simulation runs per configuration for experiments is the determination of confidence intervals. Since the calculation of the necessary number of runs requires not only the considered absolute error and the selected confidence interval but also the values for mean and standard deviation of the results, which are not known in advance, an iterative procedure is necessary. This involves running simulations, determining the two parameters, checking whether the confidence constraint is met, and then running more runs if necessary until the constraint is met. In addition, as previously noted, absolute error and confidence interval must be selected by the user. Thus, this procedure is relatively time-consuming. (Eley 2012, p. 29)

In addition, it must be decided whether each simulation run is compared individually with reality or the average of the simulation runs. Again, the way of the usage of the Digital Twin during operation is an important factor to consider. While considering all simulation runs individually creates high complexity and complicates the decision regarding model validity, considering only the average can lead to oversimplification. A compromise could be the use of the average plus the runs with the most extreme realizations of the KPIs of interest (e.g. highest and lowest output, lead time, or utilization).

4.2.5 General conditions of validation period

For the simulation runs to be compared with the real system behavior, the general conditions of the validation period must be transferred to the model. This includes the production plan with the product variants to be produced, the quantities, the production breaks, as well as the availability of resources, such as employees, transport systems, or preliminary products and materials, if these are not modeled themselves with stochastic distributions.

4.2.5.1 Organizational information

Most of the organizational information on the validation period under consideration, such as empty shifts, production breaks, variant changes, and occupancy levels can in most cases be read out relatively easily and without further data preparation from the IT systems or databases and fed into the simulation model as run information (data class D according to Collisi (2002)).

4.2.5.2 Exceptional events

For successful validation, it is also necessary to take into account unusual events that affect the system behavior but are not represented as normal behavior of the system in the simulation model, or occur so rarely that their occurrence in the validation simulation runs is very unlikely. In addition, even if such severe events do occur in the simulation runs, it is even less likely that they will occur at the same time as in reality. Even though the time of occurrence is negligible when comparing the final KPI values of simulation and reality, it makes a big difference when considering the KPI progression. As shown in Figure 4-4, which shows the cumulated number of produced parts (output) over time in simulation and reality, the area between the two output curves is much larger in case b) than in case c), although in both cases the same exceptional event occurred in reality as well as in the simulation run. Case a) shows the case in which an exceptional event that occurred in reality is not happening in simulation and therefore leads to a bad validation results for a simulation model that otherwise behaves similar to reality.

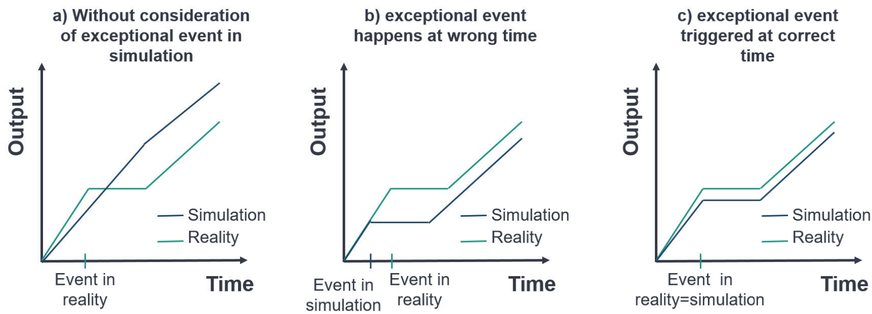


Figure 4-4 Significance of the explicit consideration of exceptional events in validation

These events, which lead to performance changes up to complete line shutdowns, can include, for example, exceptionally long failure events, due to unknown, new failures not foreseen during PLC programming, or organizational peculiarities outside the scope of the material flow simulation, such as staff meetings, trainings, or fire alarms. It must be possible to detect such exceptional events from the available operational data to perform validation in an automated manner.

The decision which events are to be considered so extraordinary that they have to be triggered explicitly in the validation simulation runs and which events are to be considered normal behavior in the simulation model comprises several aspects. On the one hand, the decision depends on the defined observation frame of the simulation model, on the other hand on the length and impact of the event. In order to be able to recognize unusual events from the available data, a differentiation according to the type of event might be possible. For this, however, these types must be contained in the data, all possible manifestations must be known in advance, and they must have been manually labeled as exceptional or normal in advance. Since these prerequisites will not always be met, or new, undeclared events may occur during operation despite extensive pre-declaration, it is recommended to implement a distinction based on the temporal length of the event instead or as a supplement rule, since the duration of an event is included in most cases during data acquisition.

The distinction based on temporal length can be done using two thresholds: a threshold t_f for the maximum duration of normal failure behavior and a threshold t_e for the minimum duration of exceptional events. Observed outages shorter than t_f are then considered normal outage behavior and used to calculate availability. Events longer than t_e are considered exceptional. These events are not used for parameter estimation but are made explicit events within the simulation when validation simulation runs are performed. To avoid ambiguity, $t_e \geq t_f$. It might be sensible to set $t_f = t_e$ as this allows all events to be categorized. If $t_f > t_e$, there may be events that require manual decisions on how to handle them because they can be labelled normal failures as well exceptional events.

There are two ways to determine the threshold values t_f and t_e: based on historical data and based on simulation experiments. Both are discussed below in general and will be implemented for the use case in section 5.2.3.4.2. t_f and t_e should be defined during the setup of the Digital Twin and not changed later for each validation run to prevent their abuse for manual improvement of validation results.

It can also be possible for failures to have different consequences for the machines. Some might lead to a 100% decrease in capacity others to only 50% decrease. This additional dimensions of failures and events will not be considered in this thesis and a complete machine stop in case of a failure is always assumed.

4.2.5.2.1 Analytical determination based on historical occurrence of events

Based on the available data, a distribution of the lengths of the registered events in the past can be created. t_f and t_e must then be selected so that a proportion as large as possible of the machine failures (in terms of number and total duration) are covered as 'normal' failure behavior (i.e. their duration is smaller than t_f). At the same time a proportion as large as possible of the marked exceptional events should be recognized as such and thus taken explicitly triggered during validation (i.e. the duration is greater than t_e).

Another analytical approach to identifying exceptional events from historic data is the use of methods for outlier detection. A simple approach would be the use of a box plot of the event durations, where points outside the whiskers will be considered exceptional events. Methods that are more sophisticated could include subsampling and ensembling, density peak clustering, or deep learning. For a recent overview of state-of-the-art outlier detection techniques please refer to Boukerche et al. (2020).

4.2.5.2.2 Empirical determination by testing limits in the simulation model

In any case, for the best choice of t_f and t_e, the candidate values should be tested during the setup of the Digital Twin in the validation and update on different validation periods and the achievable accuracy should be evaluated. This shall ensure that the Digital Twin has a sufficiently high generalization as well as can perform meaningful validations.

4.2.6 Accuracy metrics

Since the validity of a (simulation) model in general has various aspects, which were discussed in sections 2.2.7 and 3.3, the following will focus on *accuracy*, which means the deviation of the simulation model to reality, which will be quantified using so-called accuracy metrics. Which accuracy metrics should be used in the validation process depends on several factors, such as the type of production system or the strategic requirements of management. In general, accuracy metrics can be divided into four categories, which are shown in Figure 4-5. First, individual KPIs of the system such as total output, OEE (Overall Equipment Effectiveness), or scrap over the entire period under consideration can be used to compare model and reality. This is the simplest validation option and is commonly used in practice. Second, for a better temporal resolution and comparison of the dynamics in simulation and reality, KPI progression over time (e.g. cumulative system output) can be used. If the fluctuations in simulation and reality should also be included in the comparison, scatter diagrams can be used, which can be visualized e.g. using box plots, and described with characteristic values such as variance and interquartile ranges (category 3). Even more detailed validation possibilities are offered by the fourth category of accuracy metrics, which consider the states or state development of individual components of the production system, such as average utilization or utilization development of machines, buffers, or means of transport. This fourth category also includes the throughput times of individual parts or products through the system.

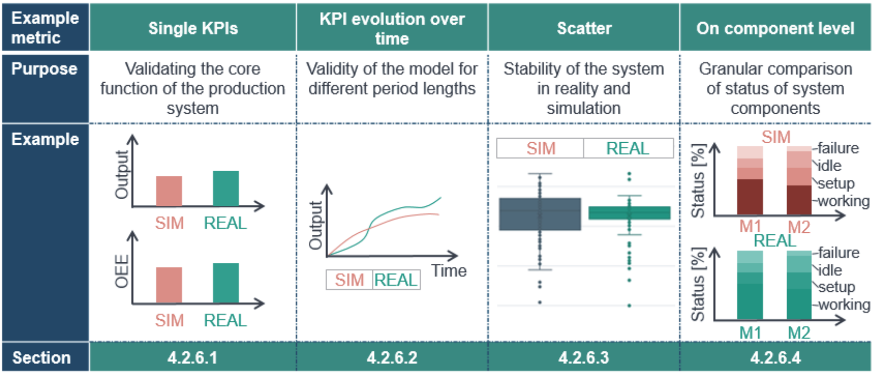


Figure 4-5 Overview of the different options for validation of the Digital Twin

In each of these categories (KPIs, evolutions, scatter, and fine granular observation), the deviation between simulation and reality can be quantified again with different metrics. These include absolute and relative differences as well as statistical test procedures (e.g. t-test).

In the following, some possibilities to measure the difference between the Digital Twin and reality will be presented. The simultaneous use of all of these metrics makes little sense due to the complexity of defining the respective threshold values and the question of how to deal with contradictory results.

4.2.6.1 Metrics using single KPI values

An easy way for a first assessment of the accuracy of the Digital Twin is to focus on its prediction capabilities of important KPIs of the system, i.e., produced parts per shift, per day, or per week. The absolute prediction error is the difference between the predicted value by the Digital Twin and the real value. Since the interpretation of the absolute error highly depends on the magnitude of the KPI itself, the relative prediction error (RE) is recommended to use for comparison of the prediction error for different KPIs, different time frames, different production systems, or different Digital Twins. Its calculation for the KPI 'produced parts' ($\mathbb{N}_{real/sim}$) is given by:

$$relative\ error = \frac{\|\mathbb{N}_{real} - \mathbb{N}_{Sim}\|}{\mathbb{N}_{real}} \tag{4.1}$$

The relative error has the value range $[0, \infty)$, often expressed as percentage and is an intuitive value to measure the difference between an output value in reality and simulation after the elapse of a predefined time period. Because it gives only a snapshot it is highly sensitive to the choice of the time period. Therefore, the next category of accuracy metrics aims to better capture the complete validation period.

4.2.6.2 Metrics using KPI evolution

To gain a first, intuitive insight into the behavior of simulation and reality over the course of the validation period under consideration, the simulated KPI curve and the real KPI curve are plotted over time in a graph as shown in Figure 4-6. This way, a first subjective assessment of the closeness to reality can be made. At the same time, an objective comparison is also made possible by calculating various prediction error metrics. The *maximum absolute error (max. AE)* quantifies the maximum difference of the KPI under consideration between the simulation and reality in the time period. This is the biggest vertical distance between the two curves at any point in time. It is calculated by:

$$max.AE = max_i(|y_i^{sim} - y_i^{real}|) \tag{4.2}$$

where y_i^{real} and $y_i^{sim}$are the values of the considered KPI at time i in reality and in simulation. Its value range is $[0, \infty)$ and it is also normally expressed as percentage. The maximum absolute error is useful with regard to automated validation since an associated limit value can be used as an upper bound for the maximal deviation between simulation and reality. This is especially important if a good agreement is achieved over the entire time period, only at one point in time or over a short time interval a high deviation occurs.

A KPI that is more robust to outliers is the *normalized root mean squared error* (NRMSE), which captures the differences in KPI development of both simulation and reality over the whole period. The NRMSE for the KPI 'parts produced' is calculated by:

$$NRMSE = \frac{1}{\overline{y}^{real}} * \sqrt{\frac{\sum_{i=1}^{N}\left(y_i^{real} - y_i^{sim}\right)^2}{\mathbb{N}_{real}}} \tag{4.3}$$

where y_i^{real} is the value of the considered KPI (i.e. 'parts produced') at time i in reality, y_i^{sim} is the value of this KPI at time i in the simulation, $\bar{y}^{real}$ is the average number of produced parts in reality (over the period), and $\mathbb{N}_{real}$ is the total number of parts produced in reality. The value range of this metric is $[0, \infty)$. (Overbeck et al. 2023)

Figure 4-6 shows an example of output curves over one week, taking into account all produced product variants and is supplemented by the accuracy metrics, which were calculated for all simulation runs.

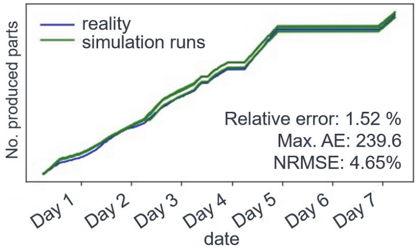


Figure 4-6 Examples of output curves

For the industrial use case of this thesis, the points at which the NRMSE is calculated are the times at which the distance between the curves changes, i.e. when either a product has been completed in reality or simulation. As shown in Figure 4-7, the time intervals between the calculation points do not have the same lengths.

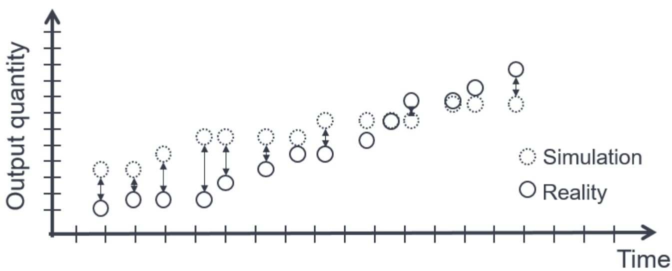


Figure 4-7 Illustration of the NRMSE calculation in the industrial use case (Overbeck et al. 2023)

4.2.6.3 Metrics using spread

Another aspect of validation is dedicated to the system volatility (and therefore KPI volatility) over time. For this purpose, the validation period is first divided into equal time intervals (e.g. of one hour). The resulting KPI values per interval are then used for a regression analysis. Time intervals in which both the simulated and real output values are zero are not considered. In addition to the coefficients of the regression lines, the coefficient of determination R^2 can provide valuable insights into the accuracy of the Digital Twin. The formulas for R^2, and the regression parameters $\hat{a}$ (intercept term) and $\hat{b}$ (slope) are given with their respective value ranges in Table 4-1. The table also includes the values that these terms would have in case of a perfect fit of simulation to reality.

Table 4-1 Overview of metrics for the validation linked to regression analysis based on Bamberg (2017, p. 40)

Metrics	Definition	Value range	Perfect value	Formula no.
R^2	$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}^{sim}{}_i - \bar{y}^{real})^2}{\sum_{i=1}^{n} y_i{}^{real} - \bar{y}^{real})^2}$	$[0,1]$	1	4.4
$\hat{a}$	$\hat{a} = \bar{y}^{real} - \hat{b}\bar{y}^{sim}$	$(-\infty, \infty)$	0	4.5
$\hat{b}$	$\hat{b} = \frac{\sum_{i=1}^{n}(y^{sim} - \bar{y}^{sim})(y_i{}^{real} - \bar{y}^{real})}{\sum_{i=1}^{n}(y^{sim} - \bar{y}^{sim})^2}$	$(-\infty, \infty)$	1	4.6

An example of the corresponding scatter plot and regression graph for one week is shown in in Figure 4-8. This graph also shows the ideal line $y^{real} = y^{sim}$ (dashed line) and the regression line (in blue). In addition, the determination coefficient and the regression coefficients are also given.

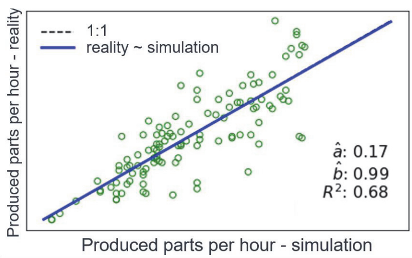


Figure 4-8 Examples of regression analysis with associated metrics

In addition to the scatter plot, the same values can also be visualized in the form of box plots as shown in Figure 4-9 for an exemplary week.

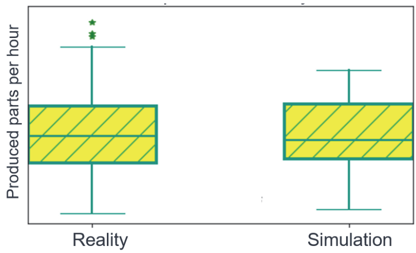


Figure 4-9 Exemplary output of the box plots

On first sight, it seems tempting to use well-known statistical tests based on hypothesis testing, for example KS- or X^2-tests, to decide whether the distributions of the model output and of the reality output can be considered to be the same. However, because the output process of both the simulation and the real system are almost certainly nonstationary and autocorrelated, these tests are not directly applicable (Law 2015, p. 269).

4.2.6.4 Metrics on component level

While the first three categories were focusing on KPIs of the production system as a whole, a good simulation model should not only be able to predict the system KPIs but also reproduce its overall behavior. This is particularly important when the model of the system is changed for experiments in order to be confident that the change will have the same effects in reality as they had in the model. To achieve this objective of an even more detailed comparison of model and reality, the behavior of each component of the system (i.e. machines, buffers, AGVs, conveyors) in simulation and in reality can be compared (i.e. utilization, states, number of parts processed).

These metrics have to be observable in reality to be of value, i.e. if the utilization of a buffer shall be used as for accuracy metrics on component level, the utilization of the real buffer has to be known at least at some points in time. These metrics therefore pose high demands on data collection. Like accuracy metrics for the whole system, they can be either calculated for certain points in time (i.e. the end of the validation period) or over the whole validation period.

4.2.6.5 Combined consideration of multiple accuracy metrics

Since the many possibilities of measuring the accuracy of the simulation model all quantify different aspects and bring their own advantages and disadvantages, which make them suitable for different scenarios, a combined consideration of several metrics might be considered. When model validity is evaluated by a human expert or a decision-making body (e.g., customers or users of the simulation model), the results of the different metrics can be weighed and a well-balanced statement can be made about the application possibilities and limitations of the simulation model. However, to close the loop of automated validation and updating, which is the prerequisite for the Digital Twin procedure, a clear, binary, automated 'yes or no '-decision is needed at the end of validation. This is possible, for example, if a weighted sum of the individual accuracy metrics is used, for which a limit value is defined. Hereby, attention must be paid to the possible value range of the metrics. Alternatively, a limit value is determined for each metric and all metrics (or a certain percentage of) must hold their threshold for the model to be valid.

The number of possible metrics to consider when performing the validation becomes especially high when metrics are calculated in component level. Then the decision whether to accept a validation result can become extremely complex.

4.2.7 Limit values for accuracy metrics

After defining the accuracy metrics under consideration for the Digital Twin, the next step is to specify the limit values above which a deviation of the Digital Twin from reality is considered too great and an update of the model becomes necessary. In addition to the naturally occurring, inherent fluctuation of the real system, the available data quality, as well as the requirements for the Digital Twin resulting from possible usage scenarios, are factors in determining the acceptable limit values. Therefore, the first step is to get an overview of the variation of the metrics in the real system over time, such as variation in output or OEE per hour, per shift, per day, per week, and so on.

4.2.8 Validation on multiple validation periods

As shown in Figure 4-10, after validation has been performed over a certain period of time and then updated with data from a corresponding period, it may be useful to test the accuracy of the Digital Twin with another validation period. This procedure corresponds to testing the model with a test data set in machine learning to prevent overfitting of the model to the given data set (in this case the given update period). Overfitting would mean that the model accurately represents the past validation and update period under consideration, but the validity for further, future periods is much lower. To avoid this, the updated model, which represents the validation period sufficiently precisely, can be tested on further, usually older validation periods.

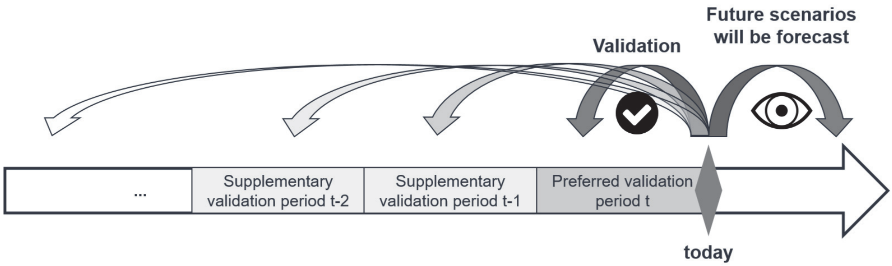


Figure 4-10 Possibility of considering several validation periods

If the validation is performed over several periods, it has to be decided how to decide if the model provides valid results for some periods but not for others. In this case, the model should be updated to provide the best possible results for the most recent validation period, even if this is at the expense of accuracy for periods further back in time. In these cases, the discrepancy can be explained by the changes in the system, and the focus should logically be on the latest period. Since these problems can arise when changes occur between the selected periods, it could be statistically verified whether the metrics under consideration (e.g., output) from all periods come from the same population. However, this is difficult to implement in practice, since the individual periods differ due to the differing conditions (see section 4.2.5) and these effects cannot always be eliminated.

4.3 Model update

If the validity of the model is classified as insufficient on the basis of the accuracy metrics, an update of the model must be carried out to improve its validity. The complete model update process, which on the one hand should enable an efficient and targeted procedure for the model update, and on the other hand, should enable the traceability of changes in the real system and of the adjustments in the simulation model is shown in Figure 4-11 and will be explained in greater detail in this section.

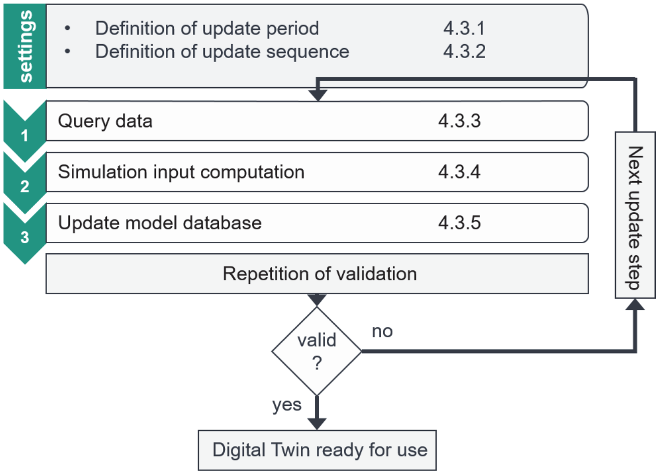


Figure 4-11 Update process and prior settings

Some settings concerning the update period and the update sequence have to be made in advance. Each update step starts with the query of the required data from the available databases (1), from which the needed inputs for the Digital Twin are calculated in the next step, which is the main challenge of the update process (2). For fast data processing it is desirable to keep the amount of data to be queried as small as possible. During input calculation the involvement of the user is possible if a completely automatic calculation does not work. The involvement of the user can range from simple yes-or-no decisions to the request to adjust the model itself. The user should be guided and supported as far as possible. Subsequently, the corresponding data in the model database is replaced (3) and a new validation is triggered to check whether the model now meets the requirements of validity.

4.3.1 Definition of update period

A crucial question in the setup of the Digital Twin is the determination of the time frame to be considered for the calculation of the new model input. For this decision opposing effects have to be weighed up: A long time period promises greater generality and reliability of results due to a larger database, but runs the risk of considering data that predate a system change and thus are representative of a system state that no longer exists. Considering such outdated data would reduce the predictive power of the Digital

Twin. A short period contains less but more recent data that may be more relevant for future forecasting. In addition, data retrieval and analysis methods run faster with fewer data. The dangers of considering a small amount of data lie in particular in the over- or underrepresentation of effects that occur with low frequency, as well as low confidence in the calculated distributions.

The trade-off must be made on a case-by-case basis and depends on the input under consideration and its fluctuations. For example, when considering the system structure, a longer period should be selected than when determining the scrap rate, as this may change between different supplier lots.

If the exact times of changes to individual machines or the entire production system are known (because they were explicitly communicated to the user of the Digital Twin or became clearly recognizable in the data after a negative validation), affected information in the next update step should only be calculated with the available data from the time after this change.

4.3.2 Definition of update sequence

The sequence in which individual components of the model are updated in the event of a negative validation has a major influence on the performance of the Digital Twin and on the traceability of its changes, on which user acceptance ultimately depends. The procedure used must therefore be documented clearly and comprehensibly.

First of all, the three information types of the Digital Twin *parameters*, *system structure,* and *dynamic behavior* have to be put in order. An intuitive approach to do so is to sort them by frequency of change, so that information that changes frequently is checked first and updated if necessary. Since parameters commonly change the most, it makes sense to address the parameter calculation first. This would be followed by the re-computation of the dynamic behavior. The checking of the system structure, which is relatively static for many production systems, would only be done at the end.

In order to determine a sequence within the parameter calculation, a sensitivity analysis, in which the components are varied individually in a realistic range with manageable step size and their influence on the considered accuracy metrics of the Digital Twin is measured, is sensible. For efficiency reasons, it makes sense to place the components with the potentially greatest influence at the beginning of the update sequence and thus update them first. As described in section 4.2.6, when using multiple accuracy metrics, a trade-off, weighting, or prioritization must be made between them, if their results behave differently with respect to the influence of individual input components. The same procedures as in the actual validation can be used for the definition of the update sequence.

Since the components of the Digital Twin can differ in their effects depending on the use case, it is not possible to define a generally applicable, optimal update sequence, but only a procedure for how this can be defined in individual cases.

4.3.3 Query data

The general structure of the data processing for the Digital Twin of the production system is shown in Figure 4-12. All data from the individual IT systems related to production are stored in the corporate data storage, which might be centralized (in a data lake) or not. The various algorithms for simulation input calculation obtain the data they require from this data storage. The calculated simulation input data is first stored in an external database and not in the simulation software itself. This corresponds to methodology C of Figure 2-7. From this external database they are then loaded into the simulation software where the simulation model can use them.

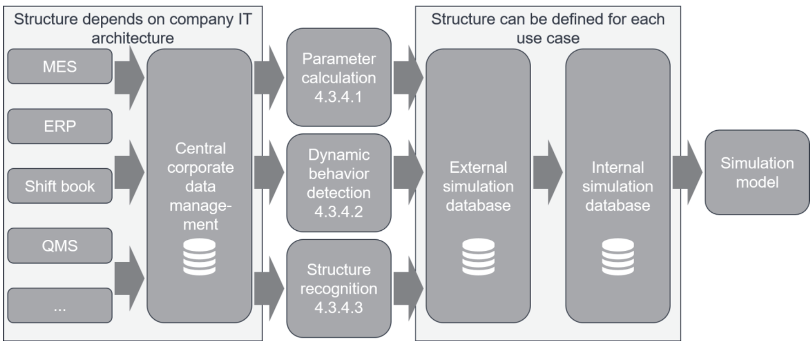


Figure 4-12 General procedure for determining the simulation input information

To compute the required input for the Digital Twin of the production system, first, the right data sources and in these data sources the right data have to be identified. This has to be done in close collaboration with the IT experts of the company as well as experts who are familiar with the production system to ensure the correct interpretation of data. If the data are stored in relational databases, SQL queries have to be defined to retrieve the necessary data in a format the algorithms for input computation are able to process. The queries can be triggered directly from the simulation model, the input computation algorithms, or a higher level Digital Twin control entity. Once the data are obtained, the simulation input computation can start.

4.3.4 Simulation input computation

A central challenge on the way to the Digital Twin of the production system is the creation of new input for the simulation model based on the company's real data.

As shown in Figure 4-13, the input required for the model update is categorized according to the classification of Collisi (2002) into *parameters, dynamic behavior,* and *system structure* (data classes P, A, and S) (see also section 2.2.5). *Parameters* are understood to be all information that can be described as individual values or statistical distributions. This includes, for example, all types of times, availabilities, or quality rates. *Dynamic behavior* comprises the rules and decisions that determine operations such as material flow, setup procedures, work flows, or assembly sequences. Approaches for the identification of material flow and work flow are presented. Information on the *structure of* the production system include not only the layout, but also information such as the number and type of machines, buffers, and transport systems, as well as their arrangements or distances from one another. As representatives of this information class approaches for the identification of used resources and work areas and walking paths are presented.

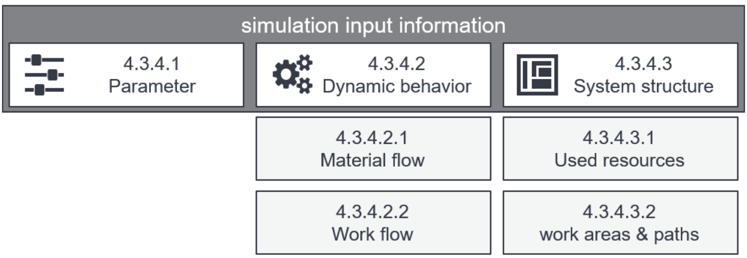


Figure 4-13 Structure of simulation input calculation

4.3.4.1 Parameter calculation

Central parameters that have to be set in material flow simulations based on real data are process times, resource availabilities, scrap rates, throughput times, and arrival processes. In stochastic simulation models parameters are described by statistical distributions from which individual values are randomly drawn during simulation runs (see section 2.2.2). These distributions must be identified from the real data and adjusted to represent reality as faithfully as possible (Law 2015, p. 280).

An alternative to the determination of continuous, stochastic distributions, which describe the historical data of a parameter as well as possible, would be the use of the historical data directly in the simulation model. Realizations of the parameter values would then no longer be generated new each time from the stochastic distribution during the simulation run, but a real, historical value would be taken. This is also referred to as using the empirical distribution, which is discrete because it contains the historic realizations of the value. Disadvantages of the direct use of the empirical distribution are:

- The effect of using different random seed values might become less transparent.
- A lower level of generalization and thus a possibly lower predictive power of the model when considering new scenarios.
- A great effort for changing parameters in experiments. Parameters of continuous distributions such as mean and standard deviation can be easily adjusted for experiments. If this shall be done for the entire historical data set, a large portion of the data points included must be precisely adjusted, which is a tedious task.
- Describing a system parameter as a statistical distribution can increase overall system understanding.
- Due to insufficient data quality, it may be necessary to clean or edit also empirical distributions before using them. This step can be accompanied by a distribution determination (see next section).

For these reasons, for the approach to Digital Twins of production systems developed in this work, statistical distributions are calculated.

Due to the often insufficient data quality in real production systems, the available raw data for distribution determination must first be prepared in most applications. Reasons for poor data quality are e.g. inaccurate data definition or insufficient coverage of special cases. A crucial step in this process is the filtering of outliers, which in the procedure for determining the distribution of machine process times presented in the following proceeds iteratively with the verification of the agreement between theoretical and historical distribution.

For the determination of the correct statistical distribution of parameters, a procedure for the supported cleaning of data (elimination of outliers) and automatic calculation of a suitable distribution was developed, whereby a normal or a lognormal distribution of the process times is assumed (A_Haizmann 2020). This can be automated by iteratively cutting off outliers and checking the distribution using statistical testing procedures (A_Merker 2020).

Assuming that the data originate from a common distribution function such as the normal distribution, the lognormal distribution, the exponential distribution, or the uniform distribution, a hypothesis test can be performed automatically for all distributions. Depending on the number of measured process times, this can be performed, for example, using the AD-test (see section 2.3.3). Input parameters for the automated distribution determination are the filtered machine process times, the test procedure to be used, the significance level α at which the null hypothesis H_0 is rejected, and the distribution functions from which the data can potentially come. As a result, the different distribution functions are returned with all parameters and the information from which distribution function the data can potentially come. Figure 4-14 shows a comparison of right-skewed data to a normal distribution, a lognormal distribution, an exponential distribution, and a uniform distribution. The green frame indicates which distribution function is a good fit to the data, based on a hypothesis test. In this example, it is the lognormal distribution. (A_Haizmann 2020)

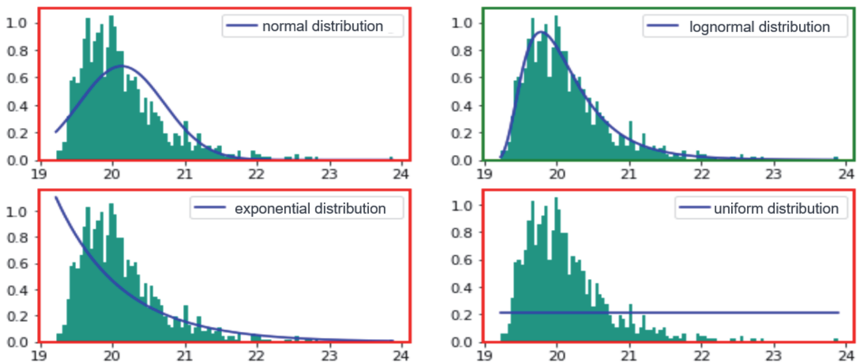


Figure 4-14 Histogram with four different distribution functions. The lognormal distribution fits best (A_Haizmann 2020)

4.3.4.2 Dynamic behavior recognition

As can be seen from the state of the art in section 3.6 the automated recognition of processes and rules of the production system for the material flow simulation is a great challenge, which has not yet been solved, due to the problems in data acquisition, recognition, representation, and forwarding to the simulation model. Therefore, in addition to existing operational data, data from localization systems are used for this task in the present approach.

Due to the variety of dynamic behavior in the production system, the material flow is considered as the central dynamic of production, which can be found in all use cases. Furthermore, a procedure for the identification of work flows is presented as a second example of important dynamic behavior in production.

4.3.4.2.1 Recognition of material flow

For the detection of the material flow, two options based on the collected operational data and an approach using further localization data collection are discussed below.

4.3.4.2.1.1 By means of process mining

One possible approach for recognizing processes in the production system from the collected operational data uses the alpha algorithm from process mining described in chapter 2.3.5.2. The procedure has already been described in Overbeck et al. (2021a) and is based on A_Teufel (2020).

Figure 4-15 Sequence of material flow detection by means of process mining

As shown in Figure 4-15, the first step is to retrieve data from the corresponding databases (1). The required data from which the event log (2) is generated should at least include information on the event number, station, workpiece position, variant produced, part ID, timestamp, and the process result (‘OK ‘ or ‘Not OK‘ (NOK)).

Before applying the alpha algorithm, the raw event data must be cleaned to identify the main flow of products to be passed to the simulation model. The following filters (3) are used for this purpose (A_Teufel 2020, pp. 47, 61):

1. Elimination of mapping information used to merge tables of different components in the assembly

2. Elimination of process chains that contain a ‘Not OK’ test result
3. Elimination of process repetitions
4. Determination of the most frequent process chain length and then exclusive consideration of process chains with this length
5. Determination of the most frequent start process and then exclusive consideration of process chains with this start process
6. Determination of the most frequent end process and then exclusive consideration of process chains with this end process

The filters listed strongly restrict the process sequence obtained to the main process. Optional and recurrent process flows are not captured in this way and must be entered manually into the model.

With the alpha algorithm (4), the relevant material flows can now be identified and represented as a Petri net. In the final transformation step, these still have to be translated into routing tables (5) that the simulation model can process (6). To create tables from the petri nets, the order of process steps is listed for each product variant and for each alternative. Due to the strict filtering, only the main process sequences are obtained which reduces the variability of process sequences to a manageable degree.

4.3.4.2.1.2 By means of machine learning

Other possibilities arise from the implicit mapping of the processes using ML. For example, an attempt can be made to learn the next control action by means of regression using a large amount of historical data. For this purpose, a data set must be prepared in which the control decision that was made in which systems status is stored, i.e. this would classify as supervised learning. Alternatively, an attempt can be made to predict the next decision using a reinforcement learning approach similar to Kuhnle et al. (2019) where the RL agent is rewarded for actions that were also taken in similar situations in reality. This allows the RL agent to learn to mimic the decision logic present in reality.

This type of approaches, which were also discussed e.g. by Bergmann (2017), offer a potentially high flexibility and general applicability to a wide variety of tasks, but come with several drawbacks, including the incomprehensibility of the control logic, which makes it difficult for users to accept the simulation model and for experimental results to be interpreted. Furthermore, the large amounts of data required as well as corresponding long training times pose challenges to the application of this approach. Further problems arise in adequately formulating the machine learning problem which requires expertise and an iterative, time-consuming approach (Kuhnle et al. 2021; Overbeck et al. 2021b). Another hurdle is the use of the learned ML models in the simulation model, which on the one hand is technically difficult to implement in many simulation programs and on the other hand can bring the simulation model into blockade situations during use and thus severely limits its runnability. Blockade situations in the simulation model can arise, for example, when employee A wants to place a part in a machine X, but this machine must first be unloaded by employee B, but this is not possible because employee A occupies the workplace in front of the machine. In reality, such situations do not arise due to the intelligent problem-solving behavior of the employees, but the simulation cannot react as flexibly. In sum, these considerations lead to the decision that the ML approach is not pursued further in this work.

4.3.4.2.1.3 By means of localization data

If products, assemblies, or individual parts can be located inside the systems by means of an indoor-localization system and the machine locations are known, the material flow can also be derived from this information. Therefore not only the moving part has to be tracked constantly, but also the area under consideration (i.e. one plant) has to be divided into sub-areas which can represent certain machines, work places, or buffers. During production, it has to be registered when a part enters one sub-area and when it leaves. The definition of sub-areas can be done manually or automatically using a clustering algorithm as shown in section 4.3.4.3.2. Advantages of this data source is its broad applicability and flexibility, its major disadvantage is that it is not yet as commonly used in industry as for example MES (Mieth et al. 2019).

4.3.4.2.2 Recognition of work flows

A major challenge in the simulation of non-fully automated production systems is the realistic representation of human behavior, which requires a lot of simulation experience, industry expertise, and observation of the employees (Greasley & Owen 2018; Zülch 2019). Target or planned work flows often exist, but their degree of compliance can vary widely. Employees can make their own decisions, taking into account a variety of information (much of which is not available electronically) and incorporating much tacit knowledge and experience. Accurately modeling these decisions, which occur repeatedly even in systems with a high degree of standardization, is time-consuming. Especially in extraordinary situations, the employees are active in shaping the system and its behavior. Therefore, the accurate depiction of their work flows ultimately determines the accuracy of the simulation model.

Moreover, employees cannot be observed directly in the electronic data, because they act independently in the system and, other than automated resources, are not controlled and regulated electronically.

At the same time, human behavior in particular changes significantly over time for several reasons. For example, the composition of the team working in the production system often changes over time due to the normal employee turnover of any company. In addition, employees are capable of learning and optimizing their behavior individually or as a team. Finally, employees are often the most flexible resource that can be used in the short term to try to compensate for any kind of volatility or problems. For all these reasons, it seems worthwhile to invest additional effort in the collection of data that supports the understanding of work processes. This can be done, for example, with the help of localization systems.

If data on the location of employees in the area of the production system are available, information on the dynamic behavior in the system can be derived from it. Thus, the assignment of employees to machines can be determined, also depending on the current system status, e.g. the number of employees in the line.

As can be seen in Figure 4-16, the two-dimensional positions (x and y coordinates, since the height can be neglected in this case) of an employee have to be clustered (e.g. using the DBSCAN algorithm) to identify areas where the employee stayed for a longer time. The clusters are then defined as work areas, and points that fall outside of these clusters are interpreted as walking paths or outliers caused by non production related movement of the worker. Now that each employee's work areas are known, they can be sequenced based on the temporal information in the localization data. This sequence of the employees' tasks can then be passed on to the simulation model in tabular form.

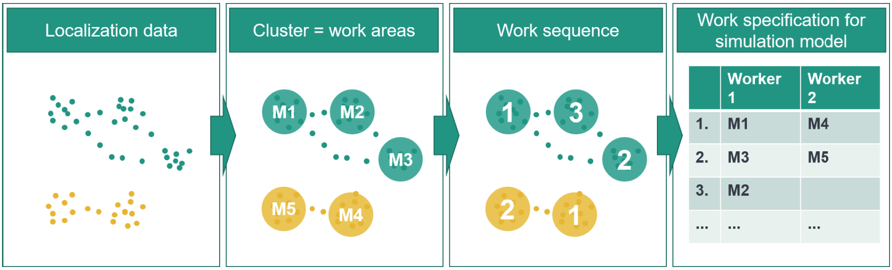


Figure 4-16 Schematic flow of the determination of the work processes in the production system

The approach presented for identifying work flows can be transferred to other processes. For example, it could be used to identify the routes of a milk run or automated guided vehicle (AGV), the use of a tool, the material flow, or a setup sequence.

The information obtained this way can also be used for structure recognition (see section 4.3.4.3.2). Under certain boundary conditions, further useful information can be obtained from the localization data, such as transport times and dwelling times, which in turn can be used to obtain indications of the duration of process steps. The problem here, however, is that it is not possible to distinguish what exactly the employee is doing, i.e. whether he is working productively on the product, carrying out preparatory activities, maintaining the machine, or simply waiting. In companies such detailed recording of employee activities is problematic due to labor protection regulations and the privacy rights of the employees. Therefore, this approach is typically not feasible in German companies.

4.3.4.3 Structure recognition of the production system

Even though the production layout in most factories is static over long periods, changes in the layout still occur in the course of its life cycle, in which, for example, machines or workstations are moved, added, or removed. In times of great overall economic uncertainty and technological change, many companies are trying to make their production systems as flexible and adaptable as possible (Fisel et al. 2019a; Fisel et al. 2019b), which in turn leads to more frequent structure changes. The Digital Twin should therefore also be able to accompany such frequent changes in the production system structure. Because these structural changes are nevertheless still carried out rather infrequently, the period between the validation and update cycles can be chosen longer than for the other two information types. Figure 4-17 shows some important considerations when changes in the structure of the production system appear. The number (1), position (2), or shape of the assets of the production system (i.e. machines, buffers, equipment, and conveyor) can change. Changes in the shape of assets might be imported directly from CAD data if it is available, but have often no influence on the function of the production system. Changes in the structure cause changes in parameters or the parameter of the new machine have to be defined (3) and changes in dynamic behavior (4), e.g. what are pre- and successor machines. Therefore, after an identified structure change further updates have to be triggered.

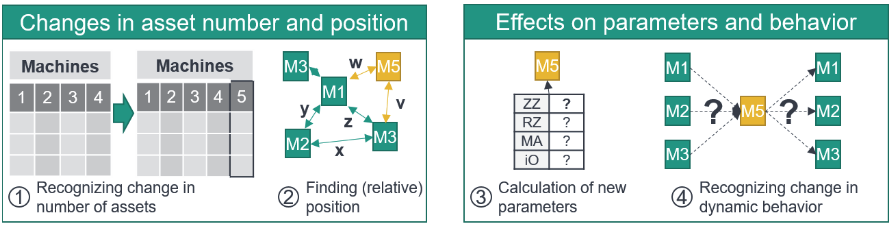


Figure 4-17 Considerations of changes in the production system

To be able to recognize changes in the structure of the production system, several possibilities will be discussed in the following.

4.3.4.3.1 Identification of used resources

Since layout information is often only available in poorly automatically readable and hardly dynamically adaptable file formats such as .jpg, .pdf, or even .pptx, and frequently represent planning statuses that have not always been adapted to reality, an attempt is made using real data from the production system. Using production data acquisition, the first part ever produced on a machine with the associated timestamp can be recorded in the same way as the last part (until today) produced on a machine. Thus, it can be determined whether a new machine has been added in the last X months or whether a machine has stopped producing in the last X months.

With the selected procedure, it can be quickly detected whether new machines have been commissioned in production. As shown in Figure 4-18 the machine can be transferred to the simulation model as soon as the first product has been produced there (even if the machine parameters such as scrap, availability, and process time are probably still insufficiently accurate at the beginning due to the initially small database).

When a machine is taken out of operation, reliable detection and, in particular, determination of the time of decommissioning is hardly possible. A machine can still be available, even if it has not been used for a long time. In the presented approach, therefore, a limit value must be set for how long a machine should not have been used before it is defined as no longer active in the simulation model (e.g. one year without a produced part on the machine). This period can, for example, correspond to the downtime period after which repeated tests and thus a longer restart is necessary when using the machine again. These additional tests and/or commissioning steps might be necessary for quality reasons or because of customer regulations. However, even with such a rule, errors cannot be excluded, so it should only trigger a hint to the user to compare the structure of the real system with the simulation model and not remove machines automatically from the simulation model. Furthermore, it should be considered that an existing but unused machine in the production system usually has no great influence on the system's behavior.

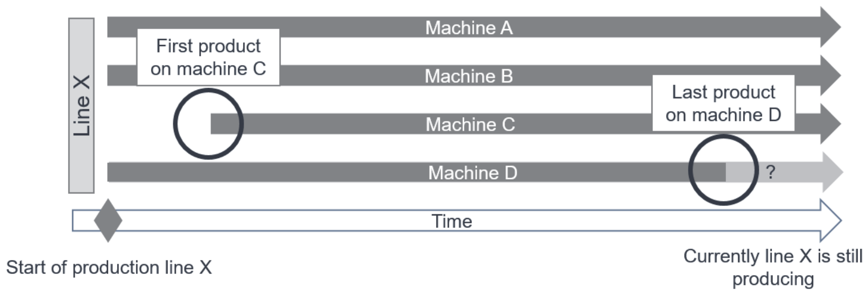


Figure 4-18 Schematic detection of commissioning and end of operation of machines A-D in line X

If the material flow is recognized automatically from the available data with the use of process mining as described in section 4.3.4.2.1.1, information about the structure of the production system can also be derived from it, because all information about used machines are included in the material flow data. The process mining approach therefore does not only provide insights into the dynamic behavior of the system but also allows conclusion about its structure. Furthermore, if the resolution of the data is sufficient, information can be obtained not only on the level of machines, but potentially also on the tool positions, workpiece positions, etc. within a machine.

However, it is important to adjust the filter settings of the process mining, because for deriving the main material flow relatively rare process can be suppressed. However, since for the layout check the focus is no longer only on the dominant material flow, but on the entirety of the line, the view must be broadened and more historical process flows must be taken into account. Results of the analysis of the production data for changes in the system structure can be found in section 5.3.3.3.

4.3.4.3.2 Identification of work areas and walking paths

If a localization system is used to detect employee movement as described in section 4.3.4.2.1.3, this can also provide important insights into the structure of the production system. Thus, areas in which the employee stays for longer time can be identified as workplaces.

The use of employee localization also has the advantage that, in contrast to the previously presented options, the walking paths can be explicitly mapped and thus non-production-related obstacles in the walking and transport paths, such as columns, offices, or lounge areas, can also be detected. In addition to the improved spatial representation of the production area in the model, this might enable a detailed verification of the required walking times between machines.

If parts or material is localized inside the plant, these data can also be a valuable input for system structure recognition, as areas with longer dwelling times can either be working areas, machines, or buffer (even if they are not designated but rather improvised storage areas). If not the material itself but some kind of workpiece carrier or container is tracked, these can also be used for system recognition, but further information on when the container is empty or full would be helpful. This might be especially relevant for simulation models and Digital Twins in which the work piece or container handling plays an important role.

4.3.5 Update model database

Once all the simulation input of the current update step is computed, it has to be saved in a defined place to be accessible from the simulation model for its update as well as from users for further analysis. A systematic versioning of simulation input data is also highly recommended to make changes in the Digital Twin transparent and allow the user to refer to older versions of simulation input data if required to retrace past decisions made based on older sets of simulation input information. Old simulation inputs should not be simply overwritten by the new input information, but saved in some kind of archive. When the new simulation input information is all stored correctly, it has to be transferred into the simulation model itself. The best procedure for this step depends highly on the used simulation software, but all modern commercial simulation software packages provide some interface for data exchange.

5 Prototypical implementation and testing

The developed methodology for the creation of Digital Twins of the production system from conventional material flow simulation models is implemented in two uses cases. The first application takes place in an industrial company of the automotive supply sector, namely the Robert Bosch GmbH (which will hereinafter be referred to as ‘Bosch’). The second use case is in a laboratory environment at the learning factory Global Production of the wbk Institute of Production Science in Karlsruhe to demonstrate the use of localization data for the Digital Twin of the production system.

First, the Bosch use cases to which the developed approach previously described was applied will be introduced. Second, the implemented mechanisms for validation and update are explained. The description of the update mechanism includes the calculation of simulation input with distinction of parameters, dynamic behavior and system structure with focus on the industrial application at Bosch. Afterwards the use case learning factory is introduced to illustrate the potential of localization data for the detection of system structure and work flows.

5.1 Use case

The procedure described above is implemented and evaluated on a real industrial use case at Bosch in the Powertrain Solutions division. For a better understanding of the use case, the production system under consideration, the IT systems and simulation software used, will be described. This is followed by the description of the manually created seed model which is the starting point of the Digital Twin process.

5.1.1 Production system

The production system under consideration is used for the final assembly, testing, and completion of internal combustion engine components and is characterized by a high diversity of product variants. Production is organized according to the Chaku-Chaku principle, in which employees are responsible for loading and unloading the individual parts and products into and out of the machines, while the actual process is largely automated (Krugh et al. 2017; Zhang & Deuse 2009). In some areas of the production system, workers are also responsible for transporting the product between machines or buffers, while in other areas the transport is automated with a conveyor on which workpiece carriers move. The production system consists of an assembly area, a testing area, and a completion area. The testing and completion areas are combined in terms of organization and IT systems and are therefore jointly referred to as ‘testing ‘ in the following. The schematic representation of the subsystems is shown in Figure 5-1.

The number of employees for both assembly and testing is variable depending on the product variant and the planned production volume. However, the number of direct employees who actually work in the subsystems can fluctuate between one and the planned number due to various factors, such as illness, vacations, training, or reduced demand volume. In addition, there is an area supervisor per shift and per area, who is responsible, among other things, for technical support in the event of malfunctions, changeovers, documentation, and substitution in the line when necessary. Intralogistics employees are responsible for supplying parts to the line according to a kanban system. Intralogistics, however, are not considered in the simulation model and a sufficient material supply is assumed (see section 5.1.3).

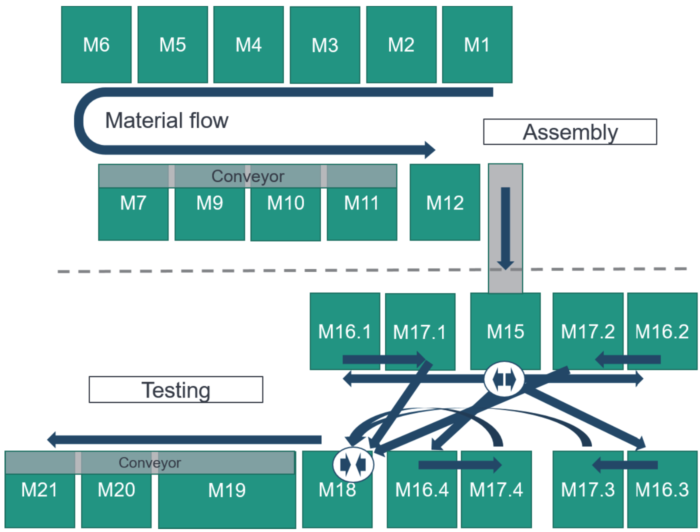


Figure 5-1 Schematic representation of the assembly and testing & completion subsystems

In the plant under consideration, there are in total four such production systems (referred to in this work as lines 1, 2, 3, and 4) for the same product, which however differ in individual machines and can thus produce different product variants. First, a Digital Twin was created for one of these lines and then transferred to the others.

Production is normally running all day long but it often happens that a line does not produce in particular shifts, e.g. due to a lack of demand, insufficiently available employees, or a lack of raw material.

Three so-called ‘sister plants ‘ abroad operate further similar production systems, but these differ not only in individual machines such as the lines within a plant but also in work instructions, product portfolio produced and used IT systems.

An important KPI used to quantify system performance is the OEE, which is calculated as the actual output of good parts divided by the target output, both per shift and per hour separately for each assembly and testing area.

$$OEE = \frac{\text{parts produced}}{\text{target parts produced}} \tag{5.1}$$

$$\text{target parts produced} = \text{target cycle time} * \text{productive time} \tag{5.2}$$

The target output for a time interval is calculated by multiplying the target cycle time by the productive time in the time interval in question. The productive time is the total time minus planned downtimes like breaks, scheduled downtimes, etc. The target cycle time in turn depends on the product variant produced, the line under consideration, and the number of employees present. If the number of worker is at the planned level, the target cycle times are normally under one minute. In the case of under-occupancy, the target cycle times are considerably higher. This means that production is possible with as little as one employee per subsystem, resulting in correspondingly long cycle times.

5.1.2 IT systems

In this use case, all data from production are collected in a MES and stored in a central data lake. By accessing this data lake, the data required for the Digital Twin of the production system can be obtained centrally. Amongst others, the data lake contains the following data in particular:

- information about each machine process instance (incl. ID and timestamp)
- test results from quality tests

In addition, target specifications are stored in an ERP system, e.g. target cycle times. Work planning is also carried out in this system, i.e. the work instructions including the so-called employee loops, which is a detailed list of tasks each worker has to perform when producing a specific product variant, and the respective target times are defined and stored in a great level of detail. As this information is highly relevant for the processes in the system, it is required, but cannot be queried automatically due to the encapsulated nature of the ERP system, which is why manual extraction of the data is necessary.

The software Tecnomatix Plant Simulation from Siemens AG[3] was selected for the simulation model in the Bosch use case, because of its widespread use in industry, the good (3D-) visualization capabilities, and the possibility to interact with others as well as the existing possibilities for adaptation through self-defined methods.

5.1.3 Seed model at Bosch use case

As described in section 4.1.2, the underlying seed model of the use case is created manually and verified and validated in discussion with experts familiar with the production system from planning and operation. Involved in the model construction were the student projects (A_Brützel 2019; A_Haizmann 2020; A_Janikovits 2020; A_Liu 2019; A_Nagel 2020; A_Xie 2020) which were supervised by the author of this thesis.

Here, the following basic modeling decisions were made:

General

- The material supply is not modeled, so the assumption is that there are always enough parts at the stations. This is consistent with observations from reality and is ensured by a kanban supply.
- The four lines are independent from each other, meaning that there are no interactions between them and they can be studied individually.
- As shown in Figure 5-2, first a model of an abstract master line is built, from which the models of the individual lines are subsequently derived (A_Nagel 2020)
- Products are fed-in at the source via a predefined production plan and run through the production system from start to finish according to the push principle. This means that in each cycle in case of an empty first station, the worker at the first station takes a new raw product from the source and begins to assemble it.

[3] Siemens AG (2023), Plant Simulation. https://plm.sw.siemens.com/de-DE/tecnomatix/products/plant-simulation-software/, [accessed on May, 5th 2023]

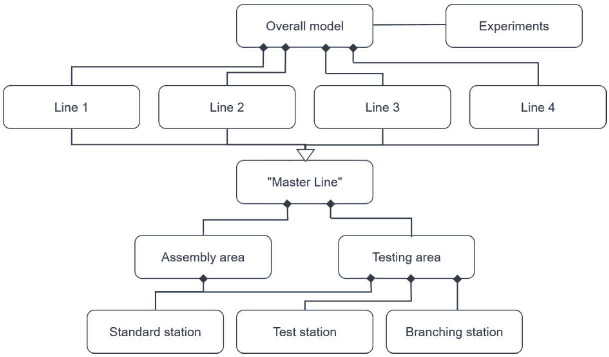


Figure 5-2 Model structure of Bosch use case based on (A_Nagel 2020)

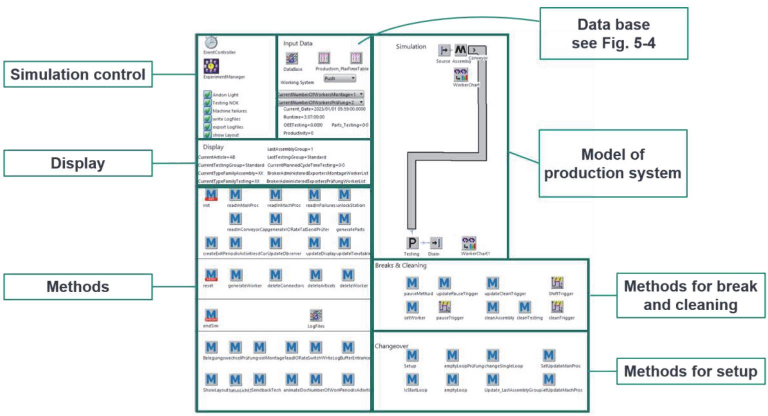


Figure 5-3 Model view in Plant Simulation (2D)

- Buffer locations and capacities are defined in a data table which is used for their creation during model initialization.
- In reality, breaks are scheduled at fixed times of the day, which are equally modeled in the simulation.
- There exist further planned and clearly defined standstill times of the machines in each shift, i.e. for machine cleaning.
- In reality, the line is never emptied during non-production shifts, except before the Sunday production interruption on Saturday evening. Therefore, the line must be refilled on Monday morning. This is handled in the same way in the simulation. Therefore, it makes sense to let simulation runs start on Monday morning so that

the ramp-up is handled equivalently in simulation and reality. With such an approach, no separate mechanism is needed to filter out any kind of warmup period.

- An internal central data storage is created in the simulation model, in which not only parameter values, but also behavioral and structural information (such as employee loops, routing table, buffers and conveyor belts) are stored in tables for each line (Figure 5-4).

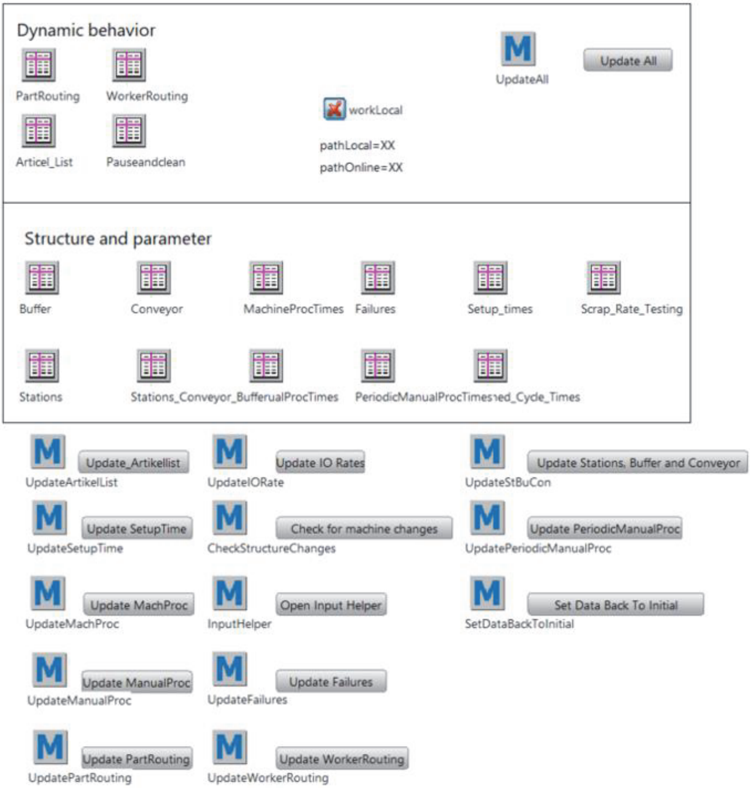


Figure 5-4 Central data management in the simulation with update methods

Stations

- Stations consist of a machine and the associated workplaces.

- All stations are structured according to Figure 5-5 with manual loading (pre), an automated machine process (machine), and manual unloading (post). Some stations require an operator to remain at the machine during the machine process, while others do not. The automated machine processes are, for example, screwdriving, joining, or testing processes. (Brützel et al. 2020)

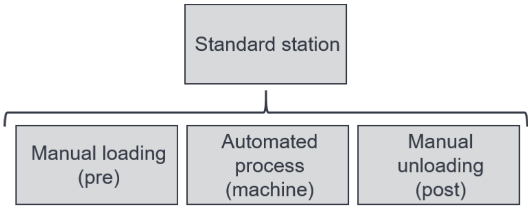


Figure 5-5 Schematic of the standard station

- Machines can have one or more workpiece positions, which enables the loading and unloading operations to be performed in parallel with the machine process or serve as an internal buffer before or after the automated process.
- The failure behavior of the machines is described in the simulation model by availability in percentage and a mean time to repair (MTTR) in seconds. The failure duration is modeled by Erlang-distributed times and the interval between failures follows an exponential distribution (Law 2015, pp. 135 & 287). When a machine is in failure mode, the employees work through their list of open tasks as long as they can. Thus, after long failures the machines and buffers after the failed station are empty and ones before the affected machine are full with parts.

Employees

- Each employee has clearly defined work instructions, i.e., a list of work steps that he must perform cyclically in each cycle on the machines in his area of responsibility, which is also referred to as a ‘loop’. This loop depends on the product variant as well as the current number of active employees in the subsystem. The loops for each worker a defined by its work instructions based on a methods-time measurement (MTM) analysis and are also visualized graphically but less exact as shown in Figure 5-6 for an exemplary product variant and four workers.

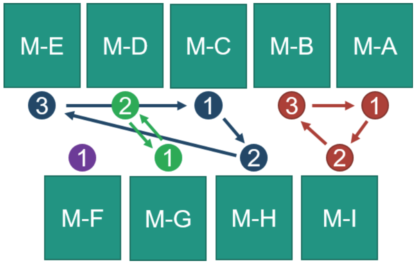


Figure 5-6 Example of employee loops in generic production area with machines A - I

- The employees receive the next task to perform from a so-called OpenRequest list (example see Table 5-1): Each area (assembly or testing) has such a list, in which it is noted which parts in which position of the line are waiting for the next action (either processing or transport) by an employee. After a worker completed a process or transport, this list is compared to the defined steps in his loop description by a central control instance, the so-called broker. The worker then executes the open request, which is next in his loop. Thus, the employee can skip steps in his loop and therefore deviate from the defined work sequence. This is necessary in order to be able to model the elimination of scrap products, the changeover between product variants, and also the start-up of the line after standstills. Slight deviations from the work specifications are also observed in reality, which is why this modeling decision is justifiable. (A_Janikovits 2020, pp. 65–68)

Table 5-1 Example of OpenRequest list (A_Janikovits 2020, p. 68)

Station	Request type	Request open
Buffer_A	Transportation	True
Station_A.Pre	Editing	False
Station_A.Pre	Transportation	False
Station_A.Machine	Transportation	False

- Handover points between worker loops can be either inside the machines (e.g. one is loading and another one is unloading after the process) or buffers between machines.
- Since the time it takes employees to perform their tasks is not recorded electronically, it cannot be observed directly in this use case. Therefore, manual process times is a non-automatable updatable input to the Digital Twin and distributions

are generated based on assumptions from planning data and estimates. For the use case, this is done by, firstly, making estimates of the maximum, minimum, and average performance of employees with respect to the planned cycle time in consultation with the work planners. Based on these times, a lognormal distribution is generated. This is particularly suitable because it can be used to approximate several successive processing steps with stochastically fluctuating times (Gudehus 2012, p. 245). In case of process changes, the new planning data for manual processes has to be updated by the simulation expert.

- In reality and the model, setup takes place machine by machine, starting from the first in line. The machines are always set up with the first part of the new variant by the shift supervisor. Since the production of a new variant, which may belong to a different assembly group, can also change the employee loops, it is necessary that each employee 'empties' his loop before he starts processing the first part of the new variant. This means that no part of the old product variant may be in any position (workpiece position in machines or buffer locations) of his previous loop, to avoid mixing of the variants or problems with the transfer of parts in the event of loop changes. This procedure does not exist in reality, but due to the complexity of the setup processes and the resulting high variability of the setup time, it does not lead to a significant distortion of the setup time on average (A_Janikovits 2020, p. 70).

Quality inspection

- There are two inspection processes that declare products as not-OK (NOK), which results in rejection of the products in simulation and reality. In reality, the rejected products go to a dedicated analysis and rework department and are usually reintroduced at different points in production (depending on the required degree of disassembly) after rework. Due to the small number of affected parts, the manifold reintroduction possibilities and the unknown duration of analysis and rework, this procedure is not included in the simulation model. Instead, rejected parts are collected in a corresponding buffer and the number of parts in this buffer is included in the results report at the end of the simulation run.
- At the beginning of the testing area, for capacity reasons, there is a splitting of the material flow into four parallel testing stations, which have the same functionality, and a subsequently dedicated post-process station for each of the four inspection stations. Two test stations are operated by one employee at a time,

which is why the material flow decision for each product at which inspection station it is tested, is not made until the branching station, depending on the free employee and the currently free test station.

An overview of the simulation model of a single line in the Plant Simulation software is shown in Figure 5-7 while the overall structure with all four lines is given in Figure 5-8.

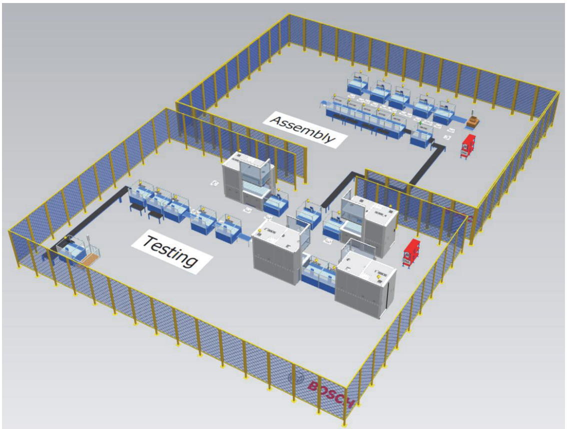


Figure 5-7 Single line in the software Plant Simulation

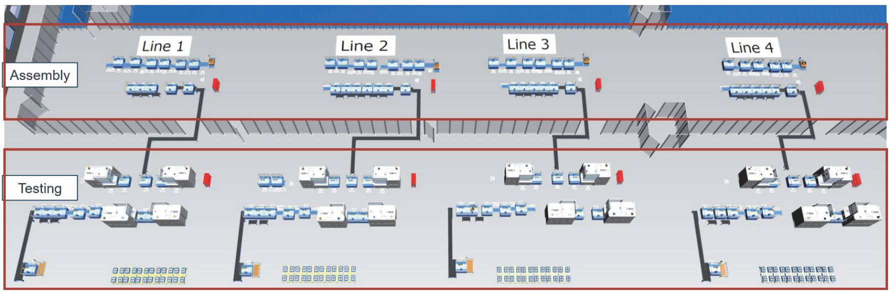


Figure 5-8 Overview of all four production lines in Plant Simulation

5.2 Model validation

For the implementation of the validation procedure described in section 4.2, the calculation of the accuracy metrics and the consideration of validation information must be specified and implemented for the use cases. The decisions made for the Bosch use case are discussed and justified in the following.

5.2.1 Implementation of validation process

The implemented validation algorithm is shown in Figure 5-9. It is indicated which process steps are executed in the Plant Simulation software and which are executed in separate Python[4] scripts. The validation process can either be started manually or automatically after a predefined interval (e.g. one week or one month).

1. First, the required data for setting the validation simulation runs are retrieved from the company database (see section 5.2.3).
2. Next, the script collects the result data from simulation and reality. For this purpose, the corresponding simulation runs are triggered and the output data of the reality are retrieved from the data lake.
3. The result data from simulation and reality are compared with each other after they have been converted into a uniform format, the accuracy metrics described in section 5.2.4 are calculated and the corresponding graphics are created.
4. If the specified limits of the accuracy metrics are violated, the update of the model is triggered. For this purpose, the required data are retrieved from the data lake and the information relevant for the model update is calculated as described in section 5.3.3.

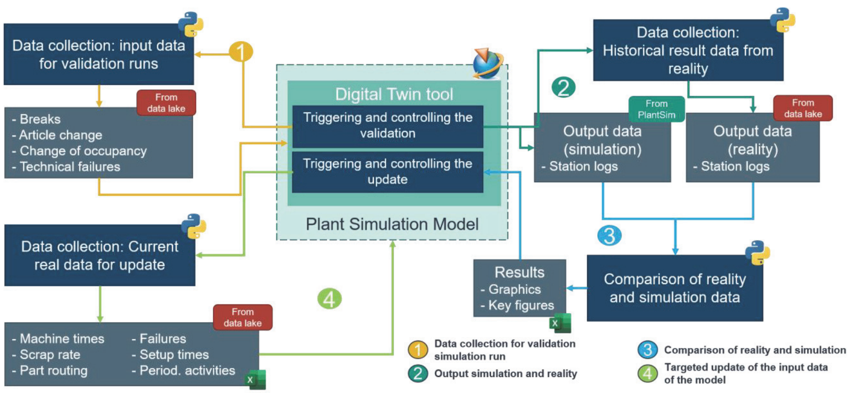


Figure 5-9 Sequence of validation based on A_Le Louarn (2021)

---

[4] Python Software Foundation (2023). https://www.python.org/ [accessed on June, 4th 2023]

5.2.2 Definition of the validation period

In the use case, the time period considered in the validation is set to one calendar week, namely from Monday, 6:00 a.m., to the next Monday, 5:59 a.m., since this corresponds to the time horizon used by internal logistics planning and also used by production planning for capacity planning, reporting, etc. Within one week, usually, a representative portfolio of products is produced and the period is sufficiently long to give a representative picture of occurring failures and changes in the number of employees.

5.2.3 General conditions of validation period

To have a meaningful comparison of the simulated week with the real week, the general conditions of the week to be simulated must be handed over to the simulation (Step 1 in Figure 5-9). In the use case, these are planned empty shifts or production breaks, exceptional events, and the number of employees.

5.2.3.1 Planned empty shifts

Empty shifts, i.e. shifts with no scheduled production within an otherwise normal production week, can occur for various reasons, such as lack of demand, lack of preliminary products, or an insufficient amount of employees available. In addition, there are production interruptions in which there is no production for longer periods not caused by a machine failure. The start times as well as the duration of these organizational standstills can be automatically obtained from the historical data in the data lake and entered in the simulation schedule for the validation week.

5.2.3.2 Product variant changeover

In the Bosch use case, there is a detailed production plan for each week, which is set up for the first time two weeks beforehand and then iteratively adjusted until the start of the week to react flexibly to external circumstances. Minimum batch size is 192 pieces. However, short-term changes in the planned production schedule continue to occur throughout the week, particularly amplified in the last few years, which were considered in this work, due to Corona pandemic and subsequent dislocations in global supply chains. Therefore, for the validation, which is always done ex-post for a given period, not the originally planned production plan for that period is passed to the simulation model, but the actual production plan that was executed (i.e. the real timestamps when

the production of another product variant was started). The transferred information includes the new product variant plus the starting time of the first part of this number at the first machine in the line. From this time on, the stepwise setup at the first station begins (as described in section 5.1.3).
This information can also be obtained automatically from the central data storage and is transferred to the source in the simulation model as a production plan.

5.2.3.3 Number of employees in the areas

Since the productivity of the line, its occupancy, and its behavior depend decisively on the number of deployed employees, this information must also be transferred to the simulation model before the validation runs. This includes the occupancy of both subsystems at the beginning of the validation period, the times when the occupancy changes, and the new number of employees in the subsystem. As described, the number of employees per subsystem is variable and used to scale the system output to current market demand.

This information is documented digitally by the shift supervisor in dedicated software, but unfortunately cannot be extracted automatically from the central data storage system. Therefore, it must be read manually from this software and transferred in a table with a defined format before validation.

5.2.3.4 Exceptional events

As explained in section 4.2.5, in addition to the planned production breaks, exceptionally long failures can also occur in the validation period under consideration, either on individual machines or on the entire line. Since such failures occur rarely, but their occurrence has a significant influence on the behavior of the line in the validation period, they must be transferred to the simulation during validation. Since in the simulation a certain historical failure behavior is already integrated through the machine availability, which was calculated from real data (the methodology will be elaborated further in section 5.3.3.1.2), this ‘normal’ failure behavior must be distinguished from ‘exceptional’ failures, which are explicitly triggered in the simulation model with a fixed start time and fixed duration during validation runs.

The different data sources also play a role when identifying exceptional events. In the Bosch use case, the data storage location is always the central data lake for production data, but different tables are used which are filled from different data sources:

- The machine availability stored in the simulation model is calculated from the automatic failure messages of the machine controls (PLC). The data cleaning and calculation steps required for this will be described in section 5.3.3.1.2.1.
- Exceptional events, which occur during production, are documented by the shift supervisor in the so-called E-shift book, a special software. In addition to its starting time, duration, and category, a brief written explanation is also recorded. These events can be partially reflected in the failure messages of the machines, but are only inadequately recorded by these since the affected machines are often switched off during repair so the duration of automatic failure message does not adequately reflect the real standstill duration.

For a successful validation, it is crucial to consider such exceptional events in the validation period and, at the same time, avoid double logging of failure patterns, as this could reduce the overall validity of the simulation model. In a nutshell, the problem could be summarized as follows: If many events of the validation period are explicitly triggered in the validation simulation runs for this period, the simulation model is 'forced' to follow a similar course as reality and probably obtains a higher accuracy for the considered period. At the same time, the generality of the validation statement diminishes. Therefore, for validation to be as meaningful as possible, it would be desirable to have to anchor as few exceptional events as possible in the simulation model during validation and to be able to achieve as much accuracy as possible only with the normal availability behavior of the machines. However, a complete omission of their consideration is not possible as visualized in Figure 4-4.

The three possibilities for identifying the exceptional events that originate from our own approach are presented here for the Bosch use case.

5.2.3.4.1 Differentiation based on the type of event

An obvious way to identify relevant exceptional events would be using the type of event, which is collected during data collection as described above. In the Bosch use case, events are categorized on two levels. A general distinction is made between technical and organizational events. It makes sense to consider all organizational events as exceptional events (such as trainings, employee information session, or company meetings), as these stem from origins out of the scope of the simulation model and are therefore not considered in the simulation model otherwise. The technical events can again be assigned to a variety of sub-categories by the documenting shift supervisor.

One is called ‘other’, into which many and particularly long disruptions fall. By analyzing these ‘other’ events, it became obvious that some should be considered exceptional event with regard to the simulation and other not. Precisely because the events are exceptional, it is difficult to define in advance a meaningful categorization that actually covers all possibilities. Therefore, not all technical events are considered as relevant for the simulation validation. In fact, the sub-categories of technical events can unfortunately not be used for the identification of exceptional events. Because the information on the sub-categories is not useful for the identification of relevant exceptional events, the duration of the events is used to identify exceptional events.

5.2.3.4.2 Differentation based on the duration of the event

One possibility to identify exceptional events to be considered independently of the categorization is the definition of a duration limit for the event. Exceptional events observed in the production system are then only considered in the simulation model for validation if their duration exceeds a certain limit value t_e. To prevent certain failure instances from being included both in the normal failure behavior of the machines (expressed as machine availability) and as exceptional events in the simulation, the definition of the limit value for exceptional events t_e should be in accordance with a limit value for normal failures t_f. Only machine failures with a duration shorter than t_f are considered as normal failures and used for the computation of machine availability.

5.2.3.4.2.1 Analytical determination of limit values based on historical data

To find reasonable values for both limits t_f and t_e, it makes sense to get an overview of the historical distribution of the failure durations in both the normal failure messages and the exceptional event messages. For this purpose, the failure reports of a production line over three months are given in Figure 5-10 as examples. The failure messages were evaluated according to the procedure described later in section 5.3.3.1.2.1 to eliminate pseudo-errors, pauses, etc.

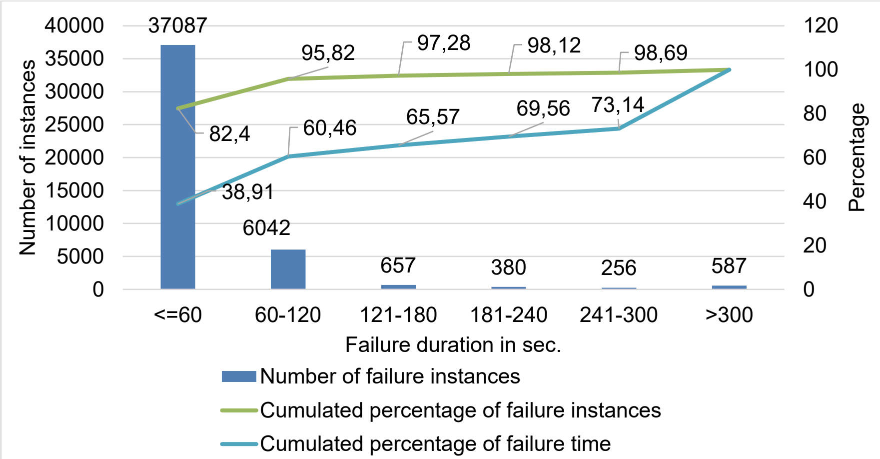


Figure 5-10 Distribution of failure durations of the automatically recorded failure messages

Supplementary information is provided in Figure 5-11, where the recorded exceptional events from four months were analyzed according to their duration.

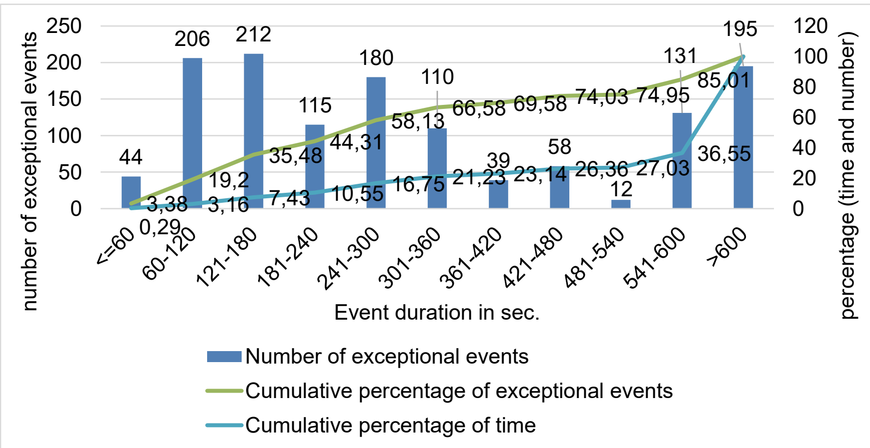


Figure 5-11 Distribution of durations of manually documented exceptional events

The analysis of Figure 5-10 and Figure 5-11 suggests that a limit value of 3 minutes (= 180 seconds) as an upper limit for normal failures and as a lower limit for exceptional events ($t_e = t_f = 180$ seconds) is promising. 99% of the failure instances, accounting for 84% of the sum of all total failures durations, are shorter than 3 minutes and 65% of exceptional events, representing 92% of the duration, are longer than 3 minutes. (Overbeck et al. 2023)

5.2.3.4.2.2 Empirical determination of limit values by validation experiments

In order to verify the effect of these limit values on the Digital Twin, an investigation of the validity of the model should be carried out over several weeks with different limit values. A full factorial experiment with all possible values for both limits is too extensive for this use case. Therefore, a fractional factorial experiment design was executed.

The resulting relative error created by update and subsequent validation with different limit values for an exemplary week is shown in Table 5-2.The resulting relative errors for combinations of different values for t_e and t_f are shown for the assembly and the testing subsystem. The table also includes the number of normal failures and of exceptional events that result from the choice of limit value t_f respectivelyt_e. It can be seen that 3 minutes perform worse as a limit value than other limit values. Especially the limit value of 15 minutes leads to a higher accuracy. The higher limit value is chosen in this case ($t_e = t_f = 15\ min$) because higher limit values mean fewer exceptional events and more normal failure behavior, which is in the model included as common machine availability. This is desirable for a simulation model because it should lead to higher generalization and, thus, better predictions of the future. This choice also satisfies the recommendation from section 4.2.5.2 that t_e and t_f should be identical. (Overbeck et al. 2023)

Table 5-2 Example of experimental investigation of possible limits for determining availability and exceptional events (EE) adapted from (Overbeck et al. 2023)

Relative error of the total output in the validation period in **assembly**		**Lower limit value for EE = t_e (in min)**				Number of failures included
		2	**3**	**...**	**15**	
Upper duration limit for normal failures = t_f (in min)	**3**	0.60	0,33	...		123
	5	1.10	0,51	...		127
	...	...	...	...	...	...
	15			...	0,0	129
Number of included EE		102	73	…	7	

Relative error of the total output in the validation period in **testing**		**Lower limit value for EE = t_e (in min)**				Number of Failures included
		2	**3**	**...**	**15**	
Upper duration limit for normal failures = t_f (in min)	**3**	3,16	3.3	...	…	123
	5	3,75	3.1	...	…	127
	...	...	...	...	...	...
	15		…	...	2.61	129
Number of included EE		102	73	…	7	

5.2.4 Accuracy metrics and limits

In the Bosch use case, the comparison between the simulation model and reality is primarily done using the number of parts produced per time interval (hour, shift, and week) of the two subsystems measured by the number of good products passing the respective last station. The subsystems are interconnected and thus influence each other, but they are organizationally separated to some degree. The consideration of both subsystems allows high-resolution analysis of the overall system behavior and can ultimately increase confidence in the validation results as well as provide a detailed understanding of the overall system. In the simulation model each finished product is registered at the last station of the two subsystems and compared to the corresponding records from the real system. These values can be used to calculate the metrics described in section 4.2.6. Resulting values and their analysis based on different degrees and forms of updates are presented and discussed in section 6.1.

As a simple accuracy metric, the relative error between simulation and reality of the total parts produced in the entire period is calculated. Since the consideration of the relative error refers only to one point in time, it has to be used with care. The cumulative output curves are more representative for the whole period. If the curves of the simulation runs lie close to the real curve, the Digital Twin can provide accurate results for any time inside this horizon. The distance between the curves can be calculated using several metrics. In particular, the NRMSE is used for this use case as defined in section 4.2.6.2.

As upper limits for a positive validation a relative error of 3% and a NRMSE of 5% were chosen by the stakeholder of the Digital Twin project and experts of the production system. A model which satisfied these requirements was considered to be sufficiently accurate for production planning and the evaluation of improvement measures. While these requirements are strictly enforced be the Digital Twin algorithm when running automatically, they can be handled more freely (e.g. as recommendations) when assessing the validation of the Digital Twin by hand based on the collectivity of metrics and graphics. Results for calculated accuracy metrics in the use case will be presented and discussed in detail in chapter 6.

5.2.5 Number of validation simulation runs

To be able to perform the validation in less than two hours (which is an arbitrary chosen limit set by the end user of the Digital Twin), the number of validation runs was set to five. Of course this value also depends on the underlying computation power of the hardware used. The number of validation runs could be increased, if better equipment was used. It can happen that not all simulation runs terminate successfully because of the large number of new situations that arise during the simulation of a wide variety of production plans and schedules in combination with always new parameter values, dynamic behavior, and system structures over long periods and with many repetitions. Therefore, the simulation runs that abort or that run the whole period but without any change in state after a certain point due to an internal blockage situation (the simulation 'freezes'), are ignored in the validation. Thus, in these cases, the validation is performed with the remaining, successful simulation runs. It should remain at least three sucessful simulation runs. If problems occur in too many validation runs a simulation expert has to be informed and the model mechanisms themselves have to be checked. Incomplete simulation runs are either detected by the debugger of the simulation software or through a comparison of the last produced part in reality and in simulation. If the last produced part in the simulation run was more than 3 hours before the last part in reality, the simulation run is considered incomplete.

For validation, the accuracy metrics are calculated for each validation run individually and saved together with the associated graphics shown to the user (see section 6.1). For the automatic validation decision, the average of the accuracy metrics over all successful terminating validation runs is used. This average must not violate the set limit values (defined in the previous section 5.2.4).

5.3 Model update

The update process to adjust the simulation model after a negative validation result or when triggered manually, was implemented as described in section 4.3, similar to the validation process, in Python (see Figure 5-9). It is important to make the implementation as independent as possible from the used simulation software. Therefore, the new input information is stored in tabular form (in this case Microsoft Excel), which can be processed by any simulation program.

For the implementation of the update in the Bosch use case, the considered update period and the update sequence will be explained in the following section. Afterwards, the developed methods for input information calculation (parameters, dynamic behavior, and structure) will be introduced for the Bosch use case. Novel methods that use in-door localization data for the recognition of system structure and dynamic behavior can only be implemented for the laboratory use case learning factory and will be presented and discussed separately in section 5.4.

5.3.1 Definition of update period

In the Bosch use case, individual update periods were selected for each information component depending on the amount of generated data, computational effort, and frequency of changes. The chosen periods are shown in Table 5-3.

Table 5-3 Periods considered for determining the updated input

Information component	Period under consideration
Machine process times	1 month
Scrap rate	1 month
Machine availability	1 month
Material flow	The last 10.000 parts of the product variant
System structure	6 months

As described in section 4.3.1, it should always be a period selected for which sufficient production data are available. If the period is too short, the sample size obtained is not sufficient and therefore not representative. If the period is too long, the amount of data becomes too large and the response time of the database grows considerably. As a default setting, the period for machine process times is set to one month. Experience has shown that one month provide a good compromise between runtime and sample size in use cases. In most cases, it is unknown whether in the selected period data concerning a certain product variant are present or not. If no data are available in the specified period, the program reports this, and another, longer or earlier period has to be selected (A_Merker 2020, p. 64).

5.3.2 Definition of update sequence

The determination of the update sequence in the Bosch use case was based on studies of the frequency of changes, resulting in the update sequence shown in Figure 5-12.

Figure 5-12 Update sequence used in Bosch use case

First, the parameters of machine process times, scrap rate, and machine availability are updated, whereby the process times are only updated for the product variants that were previously identified as invalid. Then the dynamic behavior is updated as far as possible and only in the last step the system structure is considered.

5.3.3 Simulation input computation

The next step in setting up the Digital Twin of the production system is crucial for being able to create accurate replications of reality: the development of algorithms for the computation of the input information. In the following, the main algorithms to calculate the simulation input for the use cases will be explained, sticking to the categorization of the simulation input into parameter, dynamic behavior, and system structure.

In all cases, the input information is first stored in Excel files outside the simulation model before it is imported into it. This follows the approach C of Robertson & Perera (2002) (see Figure 2-7). Input data can therefore be checked before updating the Digital Twin and it is also easier to store for documentation purposes. This approach makes the Digital Twin of the production system easier to manipulate, its updates better understandable and in general more acceptable for its potential users.

5.3.3.1 Parameter calculation

After modeling and implementing the seed model, the procedures for parameter calculation were developed. Due to the significantly larger data volumes in the Bosch use case than in the learning factory and other challenges that arise when working with real data from continuously operating systems, such as insufficient documentation and erroneous values from maintenance operations, etc., the algorithms developed for this use case are presented below.

5.3.3.1.1 Automatic determination of machine process times

The algorithm for determining the distribution of process times of machines sends SQL queries directly to the data lake, from which the stored MES data are retrieved for analysis and transformation. Outlier filtering and distribution determination are performed iteratively hand-in-hand, as described in section 4.3.4.1.

In consultation with the responsible experts for work planning, it was initially assumed that all process times of the considered machines in the production system were either normally or lognormally distributed. This assumption was confirmed in course of the analysis of the process time for each machine. To be able to determine the distribution parameters mean value and standard deviation, outliers must be removed, which can be done either manually or automatically.

In an initial analysis, the tool shown in Figure 5-13, which is based on (A_Haizmann 2020), was used to define the period under consideration, as well as the upper and lower bounds of the values to be considered. Figure 5-13 shows the transformation of process times of a machine to a probability distribution after cutting out outliers. For each machine and product variant, a distribution function is calculated. The web application visualizes and supports the distribution adjustment. On the left side, the filtered time series of parts processed at the station, color-coded as a box plot (green = between lower and upper quartile, yellow = within 1.5 times the interquartile range, red = outlier), is first displayed. On the right side, the historical process times are shown as a histogram. Above the histogram in the same graphic, the red line represents the density function of a normal distribution, which is determined from the mean and standard deviation of the remaining real data.

This manual definition of which time horizon of the data should be kept and the manual adjustment of the upper and lower limits (which corresponds to a truncation of the upper and lower outliers) allows a selection of the data to be considered. The automatically calculated parameters of the normal distribution (mean and standard deviation), as well as the number of data points included, are also displayed. The upper and lower limit have to be selected so that there is a (subjectively) good agreement between the histogram and the curve and without deleting too many data points. How many points can be deleted has to be decided by the user. Afterwards, the distribution parameters which are displayed in the center left of the image can be automatically transferred to the simulation model.

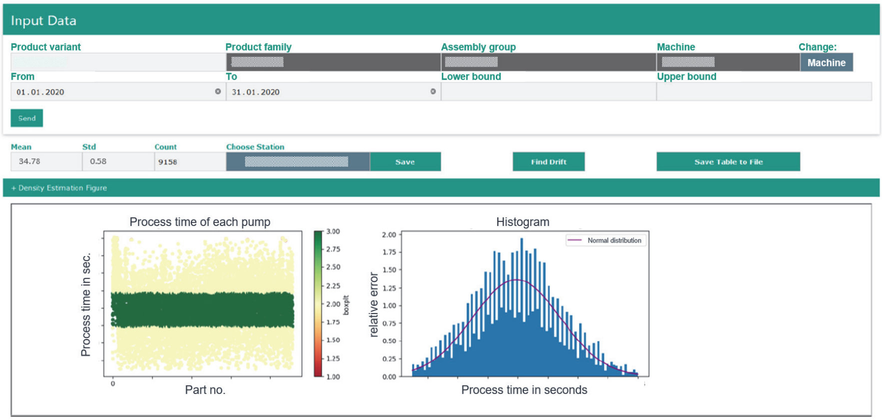


Figure 5-13 Web application with a scatter plot for a product variant at a station and fitted normal distribution according to A_Haizmann (2020, p. 77)

In order to relieve the user of the Digital Twin from manual data analysis and to enable an automated calculation of the process times of the machines, the outlier filtering was automated in a second step. The automated filtering process works without a graphical interface and interaction with the user, which makes it much faster and deterministic.

An important part to be considered in the data preparation is the handling of special cases, since the data is retrieved as raw data from the data lake. Exploring the data has shown that the data include unrealistic values, which can be excluded in advance without a detailed check by a hypothesis test or by a human, because they are, for example, negative or much too large. Thus, for the automation of this process, all negative values as well as values, that are clearly too high, (process times longer than one shift) have to be automatically filtered out from the machine process times in the step of data preprocessing (A_Merker 2020). All probabilistic distributions of process times in the simulation model are given the same lower and upper limit as used in this filter process to prevent the occurrence of unrealistic values in the simulation model as well.

After pre-processing, the data are checked for normal distribution using the AD test (see section 2.3.3). During implementation, it should be noted that the case that all recorded times are exactly the same must be handled separately. In this case, the AD test would not come to a satisfactory result at any time, so that all points would be deleted from the data set. This case can be solved by defining the occurring value as the average of the normal distribution and setting the standard deviation to zero.

To allow subsequent manual verification and possible adjustment of the filtering process and distribution calculation, the same graphs as in the manual filtering process are created during the automated input calculation and shown afterwards (see Figure 5-14). (A_Merker 2020)

Regardless of whether parameters for certain stations have to be changed or not, the complete simulation input is saved in an Excel sheet at the end. This Excel sheet is then imported into the simulation model.

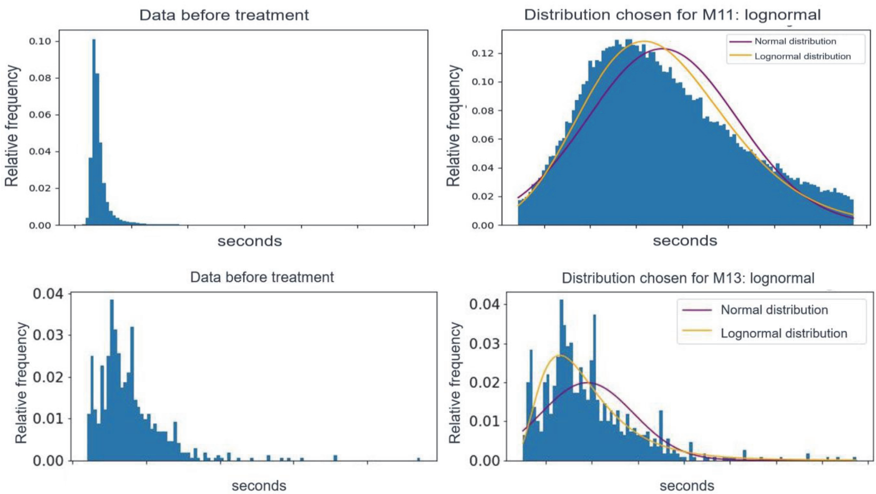


Figure 5-14 Example of distribution fitting at machines 11 and 13 for one product variant

5.3.3.1.2 Automatic determination of availabilities and failure duration

The failure behavior of the machines is determined via the two parameters availability (in %) and MTTR (in sec.). To be able to determine these two values in the Bosch use case for all machines, the machine failures must first be identified correctly. The challenge is that although failure messages from the machines are recorded electronically by the MES and stored in the data lake, the triggered failure messages cannot be equated with real machine downtimes. On the one hand, there are numerous types of failure messages that do not lead to any machine downtime. On the other hand, machines may be switched off for troubleshooting, so that the failure message is terminated, but the machine is still unable to produce. Furthermore, it can happen, for example, that a failure message is confirmed and deleted manually several times by the employee at the machine, but this does not eliminate the failure itself and the failure message keeps reappearing. This lead to several short failure messages in the MES, although in reality, it is one long failure. In addition, the duration of failure messages may be distorted by breaks, setup processes, or shift changes. The manual documentation of events by the shift supervisor in the E-shift book (already discussed in section 5.2.3.4) is a supplementary data source which has to be considered together with the MES data, because the same technical failure may appear in both systems. For all of these reasons, extensive data preparation is necessary.

5.3.3.1.2.1 Preparation of failure message data

The algorithm implemented in Python filters all stored failure messages for so-called ‘pseudo failures’ by comparing them with the production logs of the machines. As shown in Figure 5-15, failure messages during which at least two good parts were produced at the affected station are declared as ‘pseudo failures’ and ignored for the availability calculation. The threshold is not set to one good part because it might be possible that the part that is currently being processed while the failure occurs is indeed good, even in case of a real machine failure.

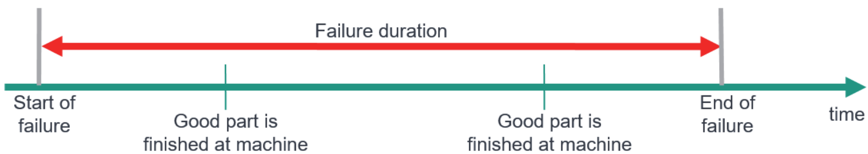


Figure 5-15 Identification of pseudo failures by comparison with machine protocol

5.3.3.1.2.2 Calculation of availability and MTTR

The availability of a machine conditioned by failure type f with failure events $s = 1, \ldots, n$ is calculated by:

$$availability_{i,f} = \frac{\text{production time} - \sum_s failure\ duration\ of\ f_s}{\text{production time}} \qquad 5.3$$

The MTTR per failure type per station is simply calculated as the arithmetic mean of the failure durations of all occurrences of this failure type at this machine in the time window under consideration.

Which failures are considered as normal failure behavior of the machine (used for the availability calculation) and which failures are so rare that they are not considered for the normal system behavior (explicitly triggered during validation), is discussed in section 5.2.3.4.

5.3.3.1.3 Automatic determination of reject rates

All test results from the test stations in the production system of the Bosch use case are stored in the data lake. The Plant Simulation software has an ODBC interface that enables a direct connection of the simulation to the data lake. The data can thus be imported directly into the data storage inside the simulation model using a SQL query. Both subsystems (assembly and testing) have their own testing station. Since the OK rate at both testing stations depends on the respective product variant, the data query is performed station- and variant-specific. (A_Haizmann 2020, p. 73)

Table 5-4 shows an exemplary result of such a query, as it is stored in the central, model-internal data storage. For confidential reasons, the numbers and yield rates are falsified. In the table, the OK rates (good yield) are defined for each product variant for each testing machine.

Table 5-4 Example results of scrap calculation using SQL from data lake

Product variant	Machine	No. of products	No. of NOK products	Good yield
A	M14	19245	498	0.9741
A	M35	19241	460	0.9761
B	M14	36452	329	0.991
B	M35	36448	1115	0.9694
...	...	...	...	...

5.3.3.2 Detection of dynamic behavior – material flow

In this work, two different aspects of dynamic behavior of the production system relevant for material flow simulation are considered: material flow and work flow. An approach to discover the material flow was implemented for the Bosch use case and an algorithm to discover the work flow using localization data will be presented using the laboratory use case learning factory in section 5.4.4.1.

The recognition of the material flow from MES data using process mining (see section 4.3.4.2.1.1) was implemented for the Bosch use case using the programming language Python and in particular the library pm4py (Berti et al. 2019). The steps presented are based on A_Teufel (2020) and have already been partially published in by Overbeck et al. (2021a).

5.3.3.2.1 Filtering material flow data

Before applying any process discovery algorithm (as for example the alpha algorithm), extensive filtering of the data is necessary to eliminate outliers. Table 5-5 shows n example of the result of such filtering on a data set of one product variant on one line over five months.

Table 5-5 Example result of the filter process as preparation for material flow derivation by means of process mining.

	All data points	Remaining share after… …removing mapping data	…removing NOK processes	…removing doubles	…filtering to most frequent process length	…filtering to most frequent start and end processes	Data used
Remaining data	26586	96%	86%	85%	85%	85%	22718

When analyzing the filter with respect to the process length for a smaller sample in Table 5-6, it can be seen that 88% of the recorded processes comprise a length of 22, which also corresponds to the maximum observed process length.

Table 5-6 Example of the distribution of number of process steps in the use case

Number of process steps	Number Process instances	Share of process instances	Share of data points
1	102	0,09	0,00
2	2	0,00	0,00
6	2	0,00	0,00
14	4	0,00	0,00
17	4	0,00	0,00
19	3	0,00	0,00
20	4	0,00	0,00
21	17	0,01	0,02
22	998	0,88	0,97

The second most frequent process length with 9% of the cases in this example comprises only one process step, which does not represent a meaningful material flow when considering the entire manufacturing process and can therefore be excluded. These tiny process sequences might be caused by setup or maintenance activities on one machine. Other process lengths occur sporadically, but are negligible for capturing the main material flow. Thus, the restriction of further analysis to processes with 22 steps is justifiable.

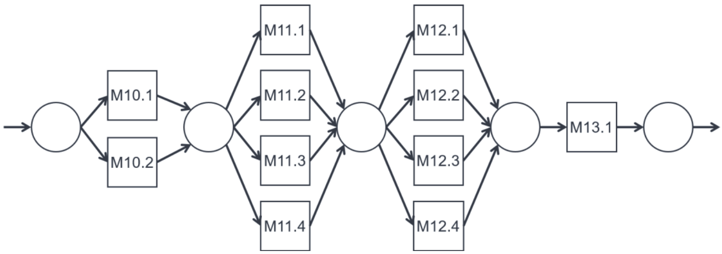
The data shown in Table 5-7 clearly shows that station 30.1 is the last station of the main material flow. For the start processes, a somewhat more complex picture is revealed: there are 4 start processes worth mentioning, where station 30.1 with a share of 7% can be excluded as a reasonable start process due to its prior definition as the endprocess. The majority of the processes (74%) start at one of the workpiece positions of Station 1, but a non-negligible proportion of the processes (17%) start at Station 2. This must be taken into account when deriving the main material flow.

Table 5-7 Example of the frequency distribution of start and end processes

Startup processes			**End processes**		
Machine	**Number of process instances**	**Share of process instances**	**Machine**	**Number of process instances**	**Share of process instances**
1.1	432	0,38	30.1	1099	0,98
1.2	415	0,36	2.1	18	0,02
2.1	195	0,17	28.2	4	0,00
30.1	84	0,07	27.3	3	0,00
12.1	4	0,00	12.6	2	0,00
11.1	4	0,00			
26.1	2	0,00			

In Figure 5-16, the effect of the filters on the complexity and variance of the material flow (here only one extract shown) becomes clear. The Petri net (a) is only a much reduced version of the complete net (b) without shortcuts and reentering flows. Version (a) can be directly used as the main material flow through the production system, whereas (b) also includes all special cases and exceptions. Since only the main material flow of the product which covers the majority of the cases is required for the simulation model, Petri net (a) is used.

a)



b)

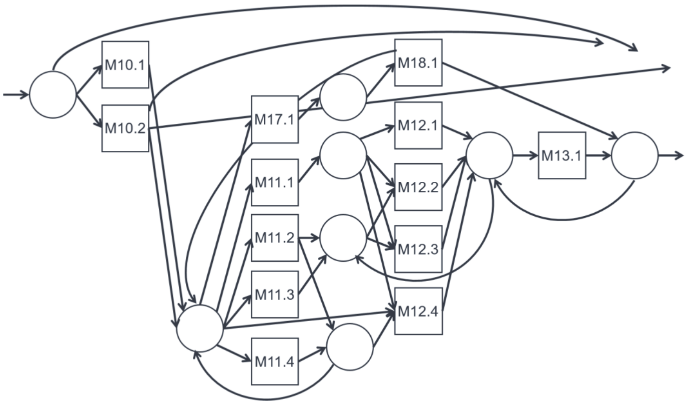


Figure 5-16 Material flow visualization with (a) and without (b) filter

5.3.3.2.2 Post-processing of the resulting Petri net

The result of the alpha algorithm can be plotted as a Petri net as shown in Figure 5-17, which shows the first section.

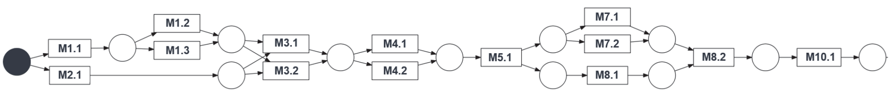


Figure 5-17 Section of the Petri net of the material flow of a product variant as a result of process mining

Since the Petri net cannot be processed by the simulation software directly, the next step is to use this Petri net to create a material flow table such as Table 5-8, which the simulation can process. Here, the sequence of workpiece positions or buffers to be reached is listed for each product variant. Each routing starts in the simulation model in the source and ends in the sink. The routings refer to the individual components of the standard station used, Pre, Machine, and Post (see section 5.1.3). The routings of the product variants can be of different lengths depending on the number of stations needed for the variant. For each process variant, only one main flow can be stored in the table. Process alternatives that are stored in the Petri net are omitted because in the use case, they only refer to different workpiece positions within a machine, but in the routing table only the machine as a whole is specified.

Table 5-8 Example routing table for the material flow of two product variants for the simulation model from the Petri net

No.	Product variant A	Product variant B	Product variant C
0	Assembly.buffer	Assembly.buffer	...
1	Assembly.M2.Pre	Assembly.M2.Pre	...
2	Assembly. M 2.Machine	Assembly.M2.Machine	
3	Assembly. M 2.Post	Assembly.M2.Post	
4	Assembly. M 4.Pre	Assembly.M3.Pre	
...	...	Assembly.M3.Machine	
...	...	Assembly.M3.Post	
48	Testing.M29.Post	Assembly.M4.Pre	
49	Sink	...	
50	-	...	
51	-	Testing.M29.Post	
52	-	Sink	

If a new product variant has to be added to the Digital Twin or the material flow of an existing variant has to be updated, a corresponding new column for its routing is created via the previously described process of filtering, alpha algorithm and transformation into a table. This column is then matched against the existing routings by the Python script. If an identical routing already exists, it used. If this is not the case, the newly discovered routing is added to the existing routing table in the central data storage.

In the course of the material flow detection, the user is provided with numerous information for a better insight into the automatic update. This includes the amount of data points at the beginning and after each filter step, the visualization of the Petri net, and analyses of process durations and variations.

5.3.3.3 Detection of system structure

With the aid of the method described in section 4.3.4.3.1, real data of the Bosch use case could be used to identify the structural changes in the production lines (which have machines of types M1-M30). The automatization of this analysis is possible by a predefined SQL query and a data processing script. Not only the actual usage period of the

Digital Twin was observed but also the historical date before the implementation of the Digital Twin to increase the number of observable changes. During the actual usage period of the Digital Twin considered for this thesis, only machine removals or decommissionings were detected, no additions (see last three years in Figure 5-18). In the earlier years, more machines were observed. Since multiple lines in one factory at Bosch are considered, some assumptions can be made about the relationships between these lines, as shown in Figure 5-18 using lines 1-3 as an example. For example, if the type 7 machine is removed on line 1 and shows up a few months later on line 3, the machine was likely moved there. Cases can also be observed where several new machines are added to a line at the same time. It is also interesting to note that the type 12 machine was added to line 3 shortly after a machine of the same type was added to line 2. It is also noticeable that machines of type 11 are removed from lines 2 and 3 at around the same time. These inferences can be used to facilitate the parameter estimation by comparing the data of the machine at the new line to the data for this machine at the old line. (Overbeck et al. 2023)

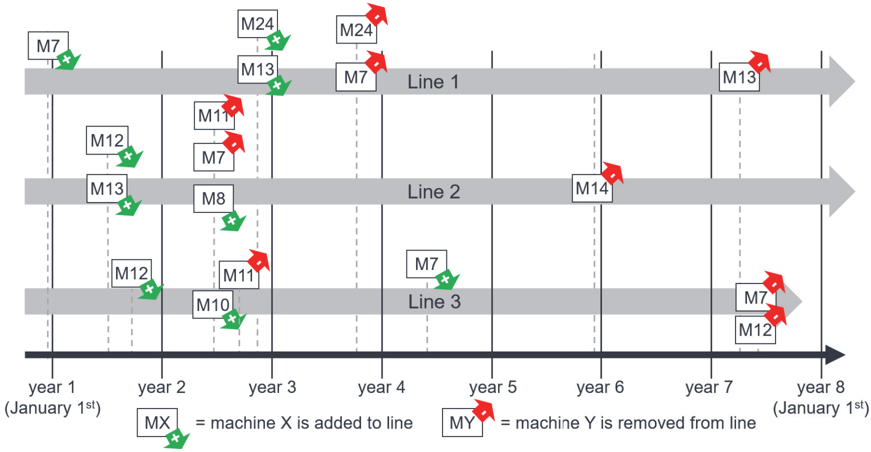
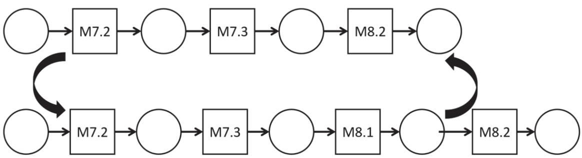


Figure 5-18 Examples of observed changes in lines 1-3 at the machine level (Overbeck et al. 2023)

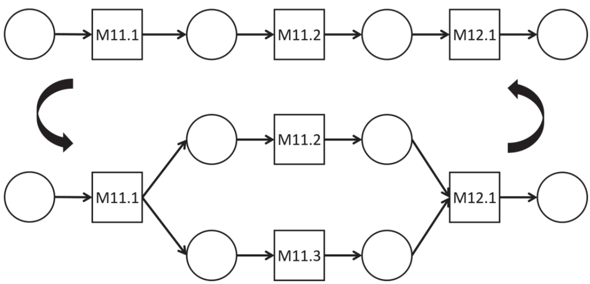
The procedure used to identify these changes in the structure of the production system was automated to be executed during the automated update of the Digital Twin. It cannot perform the required changes in the simulation model itself, because not all possible hardware changes including the corresponding layout can be parametrized beforehand. Nevertheless, the proposed approach can guide users in keeping the structure of their Digital Twin of the production system up-to-date.

Based on the detection of dynamic behavior with the help of process mining presented in the previous section 5.3.3.2., the information on the system structure could be further enhanced. Since not all changes in the processes observed from the historical data at Bosch can be listed here, some examples of types of change are shown in Figure 5-19. These are real observed changes that are used here to discuss recurring pattern. The event data used for the process mining analysis is not aggregated at the machine level but provides a higher level of detail as it includes the timestamps of the process start and end at so-called work and workpiece positions within the machines (second part of the identifier). This allows the Digital Twin to observe changes even within the machines on a great granularity. It can be noted that most of these changes do not affect complete machines, but low-level units. Changes in the number of production steps (a, b, and d (in all three examples process steps were added)), as well as changes in the arrangement of existing steps (c – the number of process steps remains constant) can be observed. When considering parallel process steps, it is important to distinguish between process alternatives (interpretable as a logical ‘or’, e.g. (d)) and cases where both parallel processes must be completed before the next process can be performed (interpretable as a logical ‘and’, e.g. (b) and (c)). (Overbeck et al. 2023)

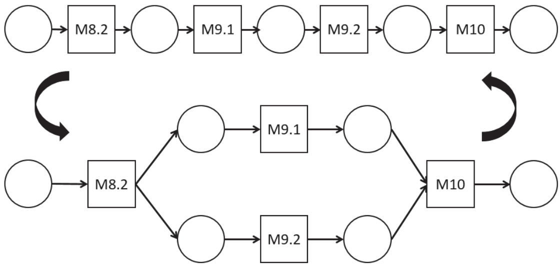
a) Adding/removing sequential steps



b) Adding/removing parallel necessary steps.



c) Switching between sequential and parallel necessary steps.



d) Adding/removing parallel, alternative steps.

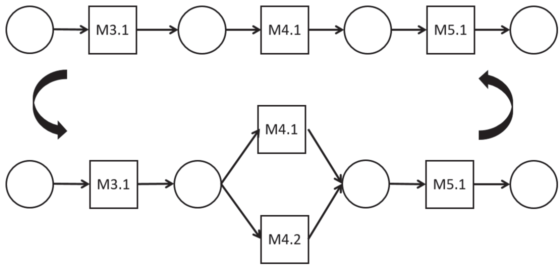


Figure 5-19 Examples of observed changes in the production process (Overbeck et al. 2023)

5.4 Model update with localization data

A second use case for the implementation and testing of the Digital Twin approach is the learning factory Global Production of the wbk Institute of Production Science in Karlsruhe. Here the developed concept can be tested in a controlled laboratory environment and the previously described use of localization technology, which is not applicable in the industrial use case, can be evaluated. A Digital Twin is created for this production system, which allows the testing of new methods for the detection of structure and dynamic behavior based on in-door localization data that are not yet applicable in the industrial use case.

5.4.1 Production system learning factory

In this production system fully functional electric motors, which are used for various applications in cars (e.g. seat adjustment), are assembled in various degrees of automation. The learning factory is a training environment for both students and professionals on topics including scalable automation, lean management, Industry 4.0, quality management, and agile production networks.[5]

The production system consists, as shown in Figure 5-20, of 10 stations, each covering different assembly process steps such as joining, pressing, screwing, magnetizing, and testing. To make the learning factory as versatile as possible, the stations have wheels and are available in several automation levels from manual to partially and even fully automated. In addition, it is possible to automate the transport between the stations with robots and/or assembly lines, so that entire production sections can be fully automated.

Figure 5-20 One configuration of the learning factory global production (wbk 2023)

The learning factory production system is extremely adaptable (especially in terms of layout and degree of automation) and thus imposes high requirements on the Digital Twin. Real production systems are usually still far away from such a degree of agility, but the applicability of the Digital Twin concept in such an extreme case thus represents a good benchmark.

[5] KIT Campus Transfer GmbH (2022), Lernfabrik Globale Produktion 2022. https://globallearningfactory.com/ [accessed on November, 23th 2022].

5.4.2 IT systems learning factory

The wbk learning factory is equipped with a MES that enables order triggering as well as picking of the required parts per order and the provision of order-specific work instructions at the workstations. In addition, it is used for data collection in the production process. The start and end times of machine processes are recorded for each order and quality problems, such as scrap, are also tracked.

In addition to the MES, there are various localization systems in the learning factory, e.g. through ultra-wideband (UWB). For this work, a system consisting of hardware from the company Kinexon GmbH and data processing software with visualization capabilities written at the institute itself is used.[6] This pre-processed data can be accessed for the Digital Twin. The hardware of the system consists of small tags, which are identified and located in their 3D position with x, y, and z coordinates, and the so-called anchors, which are permanently installed and calibrated in the hall. The tags can be attached to any object, such as machines, products, workpiece carriers, boxes, tools, or even employees, to make them localizable.

To maintain consistency with the first use case and to be able to reuse some software modules, the simulation model for this Digital Twin was also created in Tecnomatix Plant Simulation from Siemens AG.

5.4.3 Seed model learning factory

The modeling of the learning factory adopts most of the assumptions and modeling decisions of the Bosch use case. In particular, the standard station consisting of manual loading and unloading processes and automatic machining process is also adopted. Like in the Bosch simulation model, only the main part is simulated and all additional parts are considered to be always available at the right places.

5.4.4 Model update in learning factory

The laboratory use case learning factory was primarily implemented to demonstrate the potential of indoor-localization data for Digital Twins of production systems. This potential lies in the detection of system structure and dynamic behavior, both categories of information which are difficult to obtain using only MES or ERP data. Because precise

---

[6] KIT Campus Transfer GmbH (2022), Lernfabrik Globale Produktion 2022. https://globallearningfactory.com/industrie-4-0/#uwb [21.12.2022].

indoor-localization of resources and in particular worker is still used rarely in practice, this thesis resorts to a laboratory use case for the analysis of this topic. It is technically quite easy to implement but it implementation in industry is often hindered by governance or regulation issues. How localization data of worker can be used to detect the structure of and the work flows in the production system will be presented in the next two sections.

5.4.4.1 Detection of system structure using localization data

As described in section 4.3.4.3, the detection of system structure and work flows happen simultaneously when using the localization data of the worker for both purposes. Additional to the procedure described earlier, two things have to be considered when deducing and saving the locations of the workstations.

If the recording lasts long enough, dense areas may occur on the path due to frequent walking. This can lead the program to identify these dense areas as working areas and therefore create false results. To exclude this type of erroneous clusters as well as others caused by inaccuracies or other deviations, clusters containing less than p% of the total number of points are declared outliers and not considered real clusters. For the use case learning factory, p is set to 10%.

The detected clusters representing workstations with assigned machine locations are stored as convex hulls. This way, only the points that form the convex hull have to be stored and not all points of the cluster. In addition, the convex hull allows the machine locations to be viewed as a surface rather than as an accumulation of points.

5.4.4.2 Detection of dynamic behavior – work flows

The method to recognize work flows using indoor-localization systems presented in section 4.3.4.2.2 was implemented in the learning factory. In this case, the recognition of the systems structure and the recognition of work flows happen at the same time because workplaces and therefore stations are identified through the movement of the workers. The input of the algorithm is the localization information which can be visualized as shown in Figure 5-21.

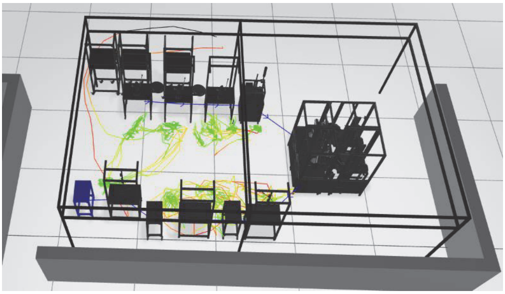


Figure 5-21 Visualization of recorded worker movements in the learning factory[7]

With the method described in section 4.3.4.2.2 following A_Kudlik (2022) the UWB localization data collected in the learning factory can be used to identify the workstations of each employee in different layouts. The clusters are determined using DBSCAN since density-based cluster analysis lends itself to the detection of frequently visited areas (see section 2.3.4). In addition, the number of clusters is automatically defined and non-spherical clusters can also be detected. The automated determination of the number of clusters is important because it may not be known beforehand or may change. For the Digital Twin the DBSCAN implementation of the *sklearn.cluster* Python library[8] is used.

The results for one setup are shown in Figure 5-22. Outliers were identified by DBSCAN and meaningful clusters were formed representing the workstations. Points in big and small clusters as well as outliers are shown in different colors. This data set includes four workers and eight stations. The machines were standing relatively close together in a U-shaped layout. Each of the employees is responsible for two workstations. The remaining stations were fully automated. The assigned stations were not necessarily next to each other. The working time was in general significantly longer than the walking time from one station to the other.

[7] KIT Campus Transfer GmbH (2022), Lernfabrik Globale Produktion 2022. https://globallearningfactory.com/industrie-4-0/#planium [accessed on June, 4th 2023]
[8] scikit-learn developers (2022), scikit-learn - Machine Learning in Python. https://scikit-learn.org/stable/index.html [accessed on June, 4th 2023]

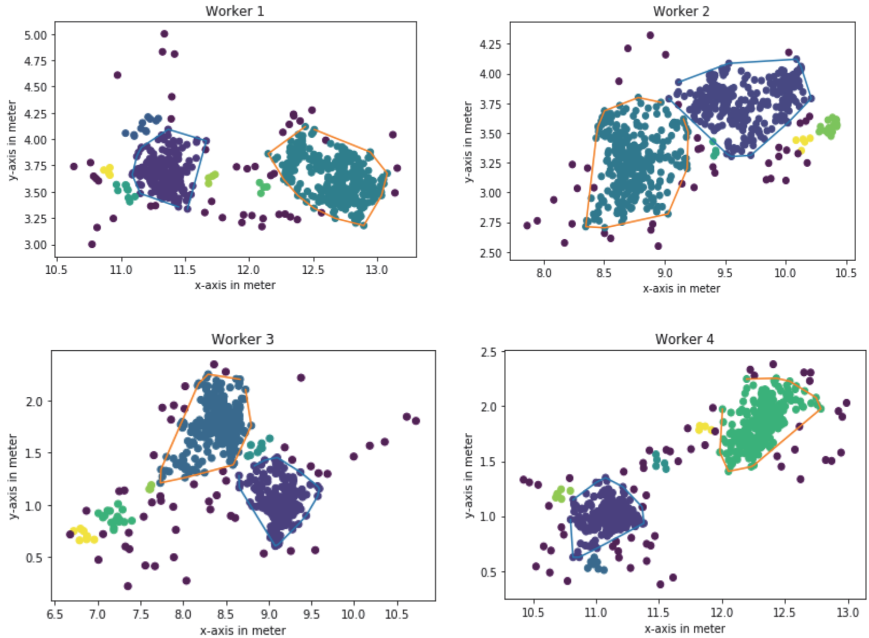


Figure 5-22 Employees' work areas identified with the aid of cluster analysis (A_Kudlik 2022)

Now, that the workstations of the individual employees are known and with the additional information on machine locations, employees can be matched with machines as shown in Table 5-9. If the timestamp of the localization signals is also considered, a work sequence can be identified. Both machine assignment and the work sequence (work flow) for each worker can be passed to the simulation model in tabular form so that worker instructions in the model are adjusted accordingly.

Figure 5-23 shows an example of the detected information employing localization information and DBSCAN.

Table 5-9 Assignment of employees to machines

Employee	Assigned machines
1	M0, M1
2	M2, M3
3	M4, M5
4	M6, M7

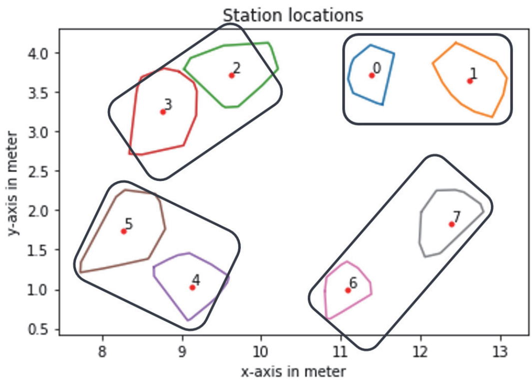


Figure 5-23 Recognized layout and employee assignment in the learning factory in the U-layout (A_Kudlik 2022)

6 Analysis of the Digital Twin of the production system

After the implementation of the Digital Twins of production systems using the developed procedures, to fulfill requirement 3 the Digital Twin of the Bosch use case will be analyzed because it incorporates all the problems that are associated with large data volumes and real life data quality. The analysis is divided into three parts: First, the degree of accuracy, which can be achieved and maintained over time using the proposed Digital Twin method, is examined. In the next step, the sensitivity of the achievable accuracy of the Digital Twin to the available data is analyzed. Finally, to demonstrate how the Digital Twin can be deployed and used in practice, a utilization concept is presented along with examples of the successful usage of the Digital Twin to demonstrate its potential for production planning, control, and improvement.

6.1 Achievable accuracy

In order to demonstrate the effectiveness of the developed and implemented procedure of this thesis for the creation of Digital Twins of production systems, first, the achievable accuracy must be demonstrated. By applying the procedure to the presented use case from Bosch, in which large amounts of data are generated, a repeated checking of the model validity is possible. The following behavior of the accuracy could be observed.

6.1.1 Initial accuracy before any update

To assess the initial accuracy of the model, a defined, executable initial state of the model is created. This is done based on an initial data set from an extended, previous period, which was used to create and initialize the seed model. This initial state of the model is validated using multiple periods with the relative error and NRMSE as accuracy metrics. The most common length of the validation period is, as justified in section 5.2.2, one calendar week, but for a more in-depth analysis of possible effects of the considered time period, further validation experiments were performed on periods of multiple weeks as well as single days. The validation is carried out on all lines in the plant considered, but for reasons of clarity, the focus in the following is on one exemplary line, where the product variety is of intermediate diversity (3 to 6 different product variants per week). In the following, this line will be referred to as line 1. The considered validation periods are shown in Figure 6-1.

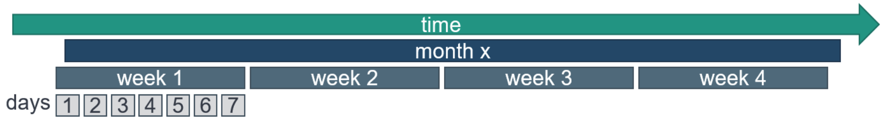


Figure 6-1 Exemplary periods for validation

Table 6-1 Initial accuracy of the Digital Twin of line 1

Validation period	Data used	Relative error (%)		NRMSE (%)	
		Assembly	Testing	Assembly	Testing
Week 1	initial	3.58	3.50	3.33	3.14
Day 1	initial	5.17	6.68	5.61	5.76
Day 2	initial	3.60	1.52	5.47	3.34
Day 3	initial	2.11	4.47	8.61	14.48
Day 4	initial	10.41	6.15	10.51	6.24
Weeks 1 & 2	initial	0.58	1.89	4.21	3.34
Weeks 1 & 2 & 3	initial	3.46	3.77	2.71	2.22

As can be seen in Table 6-1, the measured accuracy of the Digital Twin highly depends on the period considered. For single days the prediction of the simulation model might be particularly good or bad. The measured accuracy also changes when weeks are aggregated. Another important observation is that the accuracy might behave differently for each subsystem. Furthermore, it is interesting to notice that the RE and the NRMSE not always behave identical (rise or fall when changing the timeframe). This is demonstrates that they measure different aspects of model accuracy (see section 4.2.6).

6.1.2 Accuracy over time

To highlight the importance of updating with current real data, Figure 6-2 compares the accuracy measured by NRMSE over one week of the initially created model, which is regarded as a conventional simulation model with the initial data set and without updating, to the accuracy of the Digital Twin. The Digital Twin is for this analysis automatically updated every week with the latest available data. This regular update is not the default mode of the Digital Twin approach of this work, which in general includes a regular validation and only an update if necessary, to prevent unnecessary model changes. As shown in Figure 6-2, the accuracy of the conventional model deteriorates over time, while the accuracy of the Digital Twin is maintained better through the repeated updates.

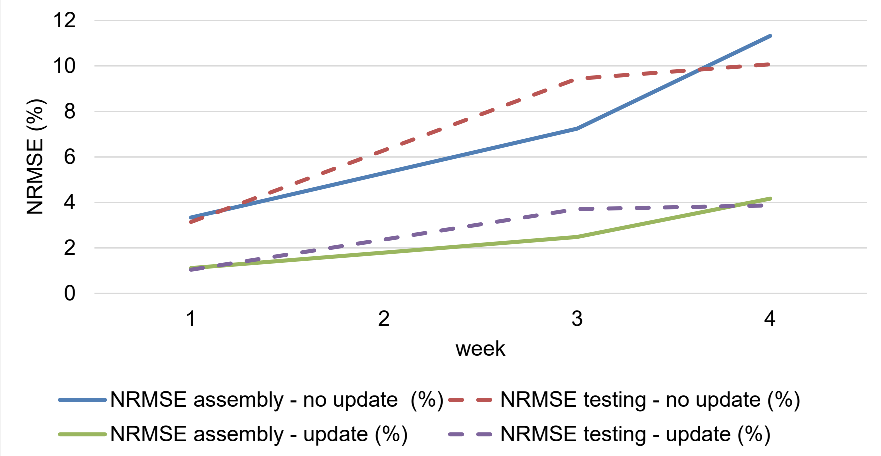


Figure 6-2 Development of the accuracy of the Digital Twin of line 1 with and without update

Because relative error and NRMSE of ‘parts produced’ are only two possible accuracy metrics, as discussed in section 4.2.6, Figures 6-3, 6-4, and 6-5 give further insights into the broad range of information and graphics that can be used to assess the validity of the Digital Twin. Figure 6-3 provides a graphical comparison of the evolution of the accumulated output between reality and validation runs for three weeks. For an in-depth analysis of system behavior it can make sense to look at only one simulation run in particular, as shown for week C in Figure 6-3. Figure 6-4 indicates that reality is stronger fluctuating than the simulation runs, but the simulation runs show a good resemblance of the average system behavior. The observation that even if the fit between reality and simulation for certain time intervals (e.g. certain hours) might be bad, the fit between simulation and reality is good in general is confirmed by Figure 6-5.

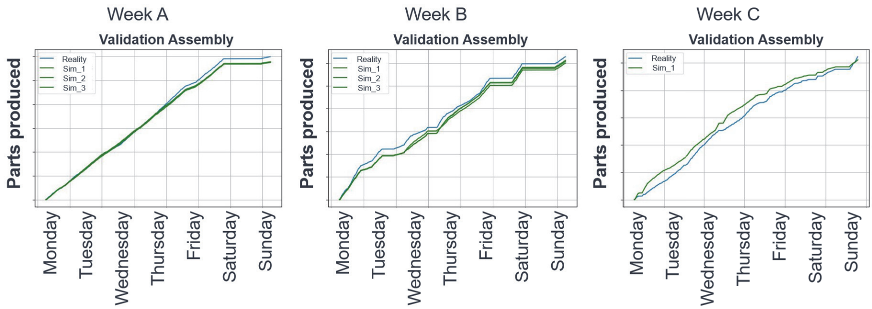


Figure 6-3 Exemplary cumulated output curves for three weeks

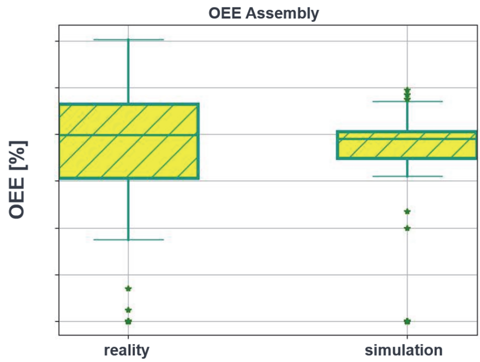


Figure 6-4 Boxplots of OEE per hour in reality and simulation

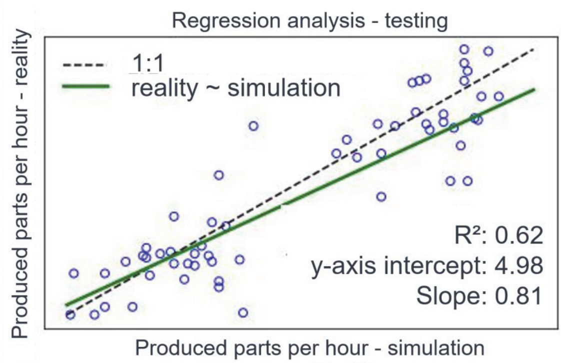


Figure 6-5 Example scatter plot and regression metrics

6.1.3 Influence of update period

To determine which data should be used to compute the new information for the Digital Twin of the production system, the influence of the so-called update period, which is the time window from which all data are considered for input computation, will be examined. Using an exemplary week, it is shown how the length of this period as well as the time between the last update and validation influence the achievable accuracy of the Digital Twin for a given validation period. Afterwards, the update period will be set equal to the validation period to create a theoretical benchmark for accuracy considerations.

6.1.3.1 Different update periods lengths

To illustrate how the validity of the Digital Twin of the production system depends on the update period, Figure 6-6 shows the achieved accuracy for different update period lengths for an exemplary validation period, which initially has a very low accuracy. 'Default' indicates that the simulation model is in its initial seed model status with the initial data set. It can be seen that the accuracy depends heavily on the period length. As described in section 5.3.1, the best accuracy achieved in this case was with an update using the data of the last month. Longer time periods provide more data but increase the risk of including obsolete data.

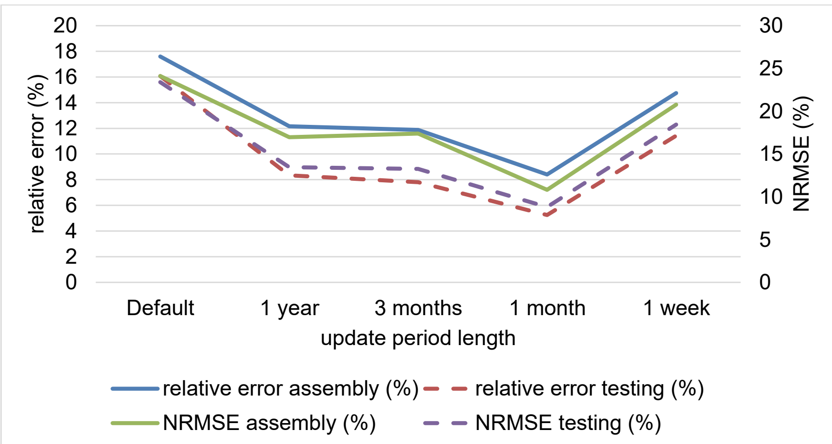


Figure 6-6 Accuracy with different update period lengths

If the update period is too short, the underlying data set is too small to make a reliable generalization and therefore the prediction capability of the Digital Twin gets worse. Because the optimal update period length depends on the use case and the information component under consideration, a similar analysis should be performed for each information component. Since it is possible that the underlying variability of input data changes, it might be advisable to repeat this analysis in longer cycles (e.g. each year), to ensure to always use the best update period length.

6.1.3.2 Time since last update

Another important factor for the accuracy besides the length of the update period is the time that has passed since the last update. To examine this, further experiments were performed with week 2. The achieved accuracy for this week by updating with data from different periods are summarized in Figure 6-7. 'Zero weeks since the last update' means that the update was performed with the newest available data, 'one week since the last update' means that the update was not performed with the data from the last week, but with data from the week before, and so on. In this example the accuracy remains stable over two weeks, gets slightlyly worse in the third week, and becomes drastically worse in week number 4. This indicates some fundamental change in the system between week 3 and 4.

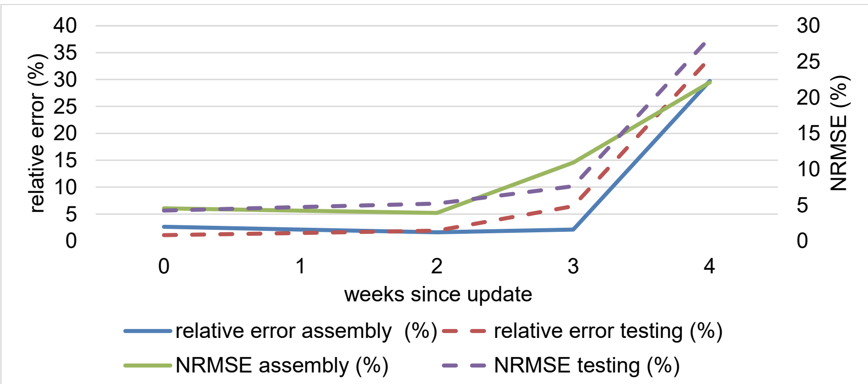


Figure 6-7 Accuracy depending on time since last update

This analysis demonstrates that a regular validation is necessary to maintain the quality of the Digital Twin of the production system. If the validation is negative, an update is triggered to increase the accuracy again. The analysis shows that the validation does not have to be repeated very frequently (e.g. every hour or day), because results do not change over such short periods but a validation every week (and the subsequent update if necessary) shall ensure that the accuracy of the Digital Twin remains high. If the validation and update cycle is performed less frequently (e.g. every month or more), the accuracy of the Digital Twin risks to deteriorate further and that the (automated or manual) update later takes much longer to reestablish accuracy.

6.1.3.3 Benchmark period

To demonstrate how the Digital Twin of the production system behaves when updating it with the actual data of the validation period, experiments were performed, in which all input is computed based on the data from the actual validation period. Thus, validation and updating period are identical. It should be noted that this is of course not possible when using the Digital Twin for forecasting, since the respective data are available only ex-post. Thus, while the forecasting capability of the model can be better evaluated by the investigations in section 6.1.2, the results shown in Table 6-2 for week 3 and 4 primarily serve as a benchmark to demonstrate the accuracy of the model with the actually realized data.

Table 6-2 Accuracy of the Digital Twin if update and validation period are identical

Validation period	Data used from…	Relative error		NRMSE	
		Assembly	Testing	Assembly	Testing
Week 3	Default	3.42	6.03	7.24	9.44
Week 3	Week 4	0.82	0.29	3.90	4.02
Week 4	Default	3.94	8.53	8.97	8.15
Week 4	Week 4	3.33	7.16	8.69	7.13

The Digital Twin is obviously more accurate when it uses data from the validation period than with default data, but this effect is not equally strong in all cases and for all accuracy metrics.

6.2 Sensitivity analysis of the Digital Twin

Many manufacturing companies cannot access all data of their machines, either because the machines are old and have no or only limited internal electronic data handling or because the company has either no right or not the know-how to access the machines controls and data storage. Another, related problem arises through the recording of wrong data points or their misinterpretation. These are just some of the many causes for low data quality in production systems. The goal of the research presented in the next section is to provide insights into which data quality and quantity is needed to create a Digital Twin of production systems with satisfying accuracy.

To address this issue, several sensitivity analyses were performed to investigate the influence of the available data on the accuracy of the Digital Twin focusing on the data quality as well as on the data completeness, as shown in Figure 6-8. First, parts of the complete real data set are removed or changed; second, the Digital Twin is updated using this altered data; third, the Digital Twin is validated; and finally the observed accuracy evaluated.

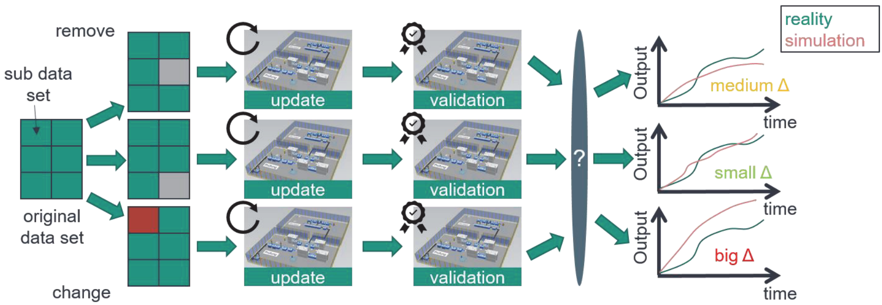


Figure 6-8 Schema of sensitivity analysis

6.2.1 Influence of data quality

To evaluate the effects of incorrectly recorded data on the achievable accuracy of the Digital Twin, the corresponding values provided to the Digital Twin are systematically varied. The focus lies initially on the information type parameter (i.e. machine process times, machine availability, and MTTR), but the importance of the correct recognition of dynamic behavior is also examined using the example of the material flow.

6.2.1.1 Machine process times

Since the machine process times are described by probability distributions, three components of this distribution can be changed: location parameter (e.g. mean), scale parameter (e.g. standard deviation), and distribution type (e.g. normal, lognormal, exponential).

It should be noted that in reality, estimation errors of these components are likely to not occur separately but at the same time, e.g. an overestimated mean (caused by outliers) may also lead to an asymmetric distribution and thus a wrongly assumed distribution type.

6.2.1.1.1 Simultaneous variation of mean and standard deviation

In a first analytical step, mean and standard deviation of the normal and lognormal distributions of the process times of all the machines are changed simultaneously and by ±% in the same direction. We vary x from -25% to +20% with step size of 5% and smaller steps around zero (±1%; ±2%). The effects on the relative error and NRMSE for the assembly and testing subsystems in an exemplary validation week including the respective 95% confidence intervals are shown in Figure 6-9 and Figure 6-10. For each configuration 10 simulation runs were performed to calculate average and confidence intervals of the respective accuracy metrics.

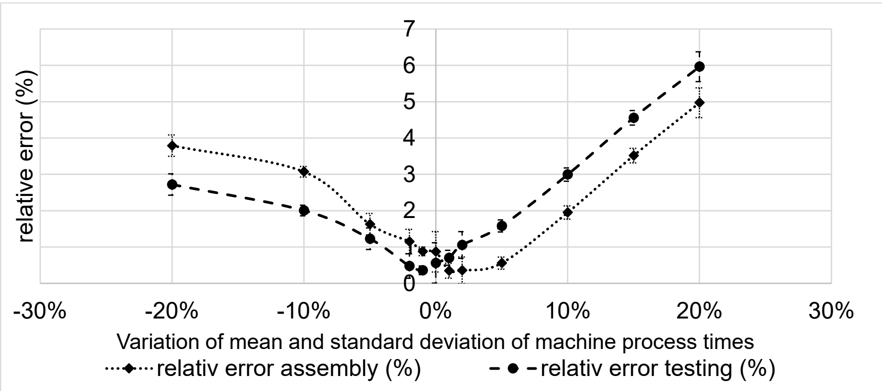


Figure 6-9 Change in relative error when varying mean and standard deviation including the 95% confidence interval

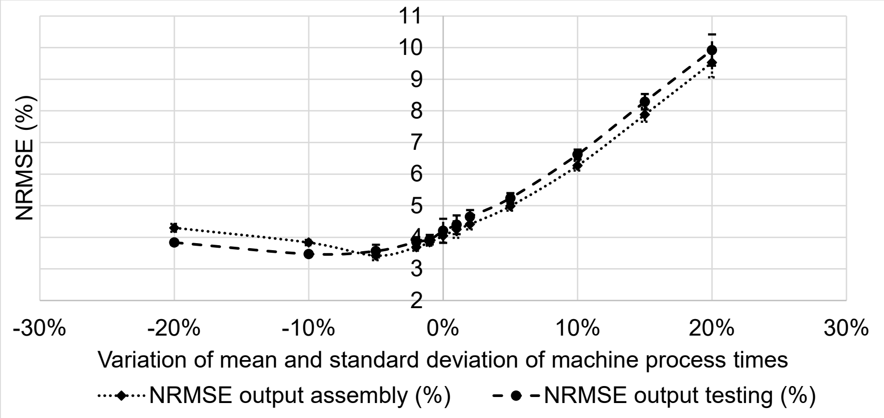


Figure 6-10 Change of NRMSE when varying mean and standard deviation including the 95% confidence interval (Overbeck et al. 2023)

The reported 95% confidence intervals for the relative error are between ± 0.12 and ± 0.56 and the 95% confidence intervals for the NRMSE value between ± 0.06 and ± 0.49. There is a small range in which a deviation of the mean of all machine process times does not have a clear impact on the accuracy and the deviations are within the normal fluctuation. The accuracy becomes significantly worse, if the estimation error for the mean machine process time recording or calculation is more than +5%. The accuracy is much less sensitive to underestimation of the parameters than to overestimation. (Overbeck et al. 2023)

The two figures also indicate that the NRMSE is less susceptible to fluctuations than the relative error. Because the relative error only considers one single point in time (the end of the period), the NRMSE proves to be a more stable and therefore, a more meaningful measure for accuracy of the digital twin since it aggregates information over the whole time period. Therefore, it is primarily considered below.

6.2.1.1.2 Variation of mean

The location of the normal and lognormal distributions assumed in the use case for the machine process times is described by the mean value of the distribution, which is relatively reduced and increased in the next step of the sensitivity analysis. This is done for all machines at the same time, since changes in process times primarily have an impact at the bottleneck process, but this may be different for other product variants.

However, due to failures and the quasi-rigid coupling of machines by employees or transport systems, process time changes at non-bottleneck machines also have an impact on the overall output. Machine process times were changed by ± 1%, 2%, 5%, 10%, 15%, and 20% in different validation weeks. For each observation, we performed 10 simulation runs. (Overbeck et al. 2023)

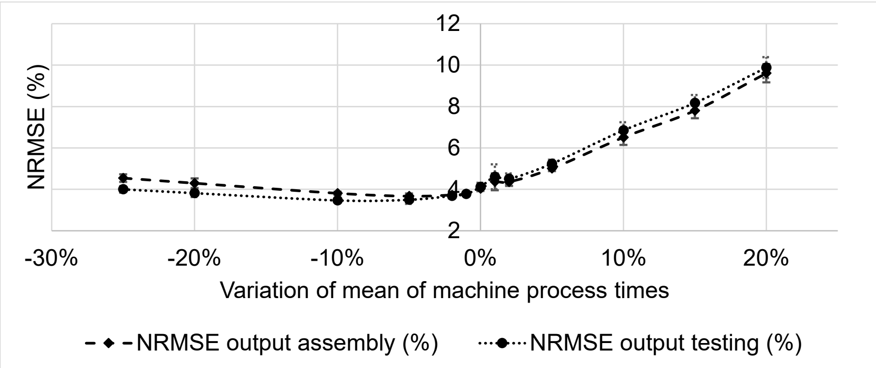


Figure 6-11 Change of average NRMSE when varying the mean value including the 95% confidence interval (Overbeck et al. 2023)

The resulting NRMSE including the 95% confidence intervals is shown in Figure 6-11. The confidence intervals lie between ± 0.06 and ± 0.6. For deviations of more than +5%, a similar behavior is observed as in the case of variation of mean and standard deviation together (Figure 6-10). The NRMSE does not continue to increase on the left side as on the right side; as the process times are reduced, at some point the worker become the system bottleneck and the system output no longer increases. (Overbeck et al. 2023)

6.2.1.1.3 Variation of standard deviation

In the distribution types used in this thesis, the dispersion is described by the standard deviation, which is reduced or increased in the steps ± 1%, 2%, 5%, 10%, 20%, 25% and 50% in the next analysis step. The mean value is thereby kept at the original, correct value. The effect of the falsification on the accuracy for an exemplary week including the 95% confidence interval is shown in Figure 6-12. Again 10 simulation runs were performed per observation. The 95% confidence intervals are between ±0.11 and

±0.47. It can be seen that the change in standard deviation does not lead to any significant change in the NRMSE. The fidelity of the digital twin seems insensitive to errors in the estimation of the standard deviation of the machine process times of up to ±50%. This means that the decrease in fidelity observed in the first analysis (section 6.2.1.1.1) can the attributed to the change in the location parameter of the distribution. Small alterations lead to a change in the system behavior but without a clear indication whether it becomes more or less accurate. This indicates that these spikes are caused by the intrinsic volatility of the system and its model. (Overbeck et al. 2023)

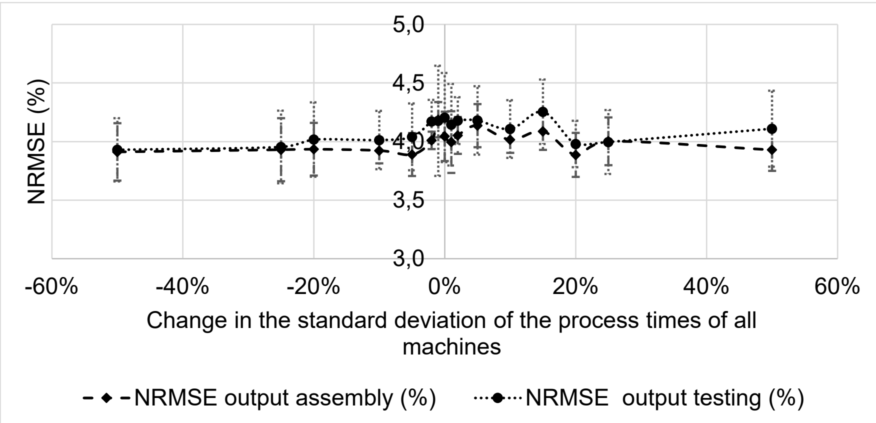


Figure 6-12 Change of average NRMSE when varying the standard deviation including the 95% confidence interval (Overbeck et al. 2023)

It can be concluded, that a deviation of less than 5% of the mean value in the data quality of the process times can be tolerated. The standard deviation of the process times is of secondary importance for the system under consideration, because high and low realizations of the process times level each other out. This result is consistent with the combined analysis of Figure 6-10 and Figure 6-11, which also indicates that the change during the simultaneous variation of mean and standard deviation is caused predominately by the change of the mean that the variation in the standard deviation has no influence in the accuracy. (Overbeck et al. 2023)

6.2.1.1.4 Variation of distribution type

Incorrectly recorded data can also result in the assumption of an unsuitable distribution type. The effects of this are also to be examined, whereby again only some exemplary distribution types, which would be conceivable for the machine process times, are considered. The corresponding parameters are estimated as good as possible from the underlying data, but errors can of course occur. For example, distribution fitting is computationally costly for distributions with a large number of adjustable parameters (Law 2015, p. 279). Figure 6-13 indicates that the choice of distribution type for the machine process times (between the considered types) is not of big importance in the Bosch use case. 'Default' describes the configuration in which the distribution type (normal or lognormal) is defined for each machine and each product variant individually based on what results in the best fit.

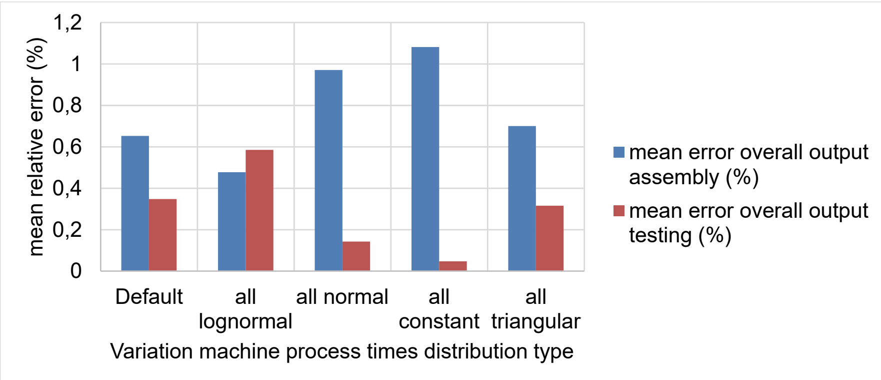


Figure 6-13 Change of relative error when changing distribution type

6.2.1.2 Machine availabilities and MTTR

Accurate measurement of machine availability is not trivial, since most machines are not always in use, but repeatedly have short downtimes during ongoing production (see section 5.3.3.1.2.1 for a discussion of the necessary data preparation steps).

In addition, there is the question of how to classify planned shutdowns for e.g. maintenance work. Even if there are clear rules in most companies whether or when they are included in the calculation of availability, these can translate differently to the simulation model depending on the selected modeling approach. The goal should be to take into account all events that lead to machine downtime, because otherwise the simulation would overestimate the capacity of the production system.

6.2.1.2.1 Variation of availability

Like the machine process times, the availability percentage of all machines is systematically increased and decreased. The underlying distribution type (exponential distribution) is not changed, as it is established for the modeling of failure behavior (Gutenschwager et al. 2017, p. 137). As Figure 6-14 indicates, an underestimation of machine availability rapidly leads to lower accuracy of the digital twin. On the other hand, an overestimation of machine availability does not lead to a higher NRMSE. This can be explained by the high machine availability of the production system under consideration. When increasing availability by only +2%, most machines are already at 100% availability (which obviously is the maximum value) and further increases have no effect. When increasing availability by +5%, all machines have an availability of 100%, so a further increase is impossible.

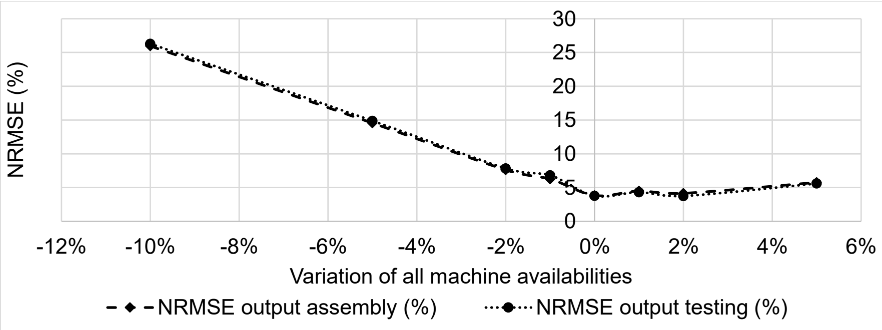


Figure 6-14 NRMSE when machine availabilities are varied

Therefore, it can be concluded, that in production system with high availability an overestimation does not significantly decrease the model quality, but an underestimation has strong effects. It is important to mention, that this can be different in systems with a lower machine availability.

6.2.1.2.2 Variation of MTTR

If availability remains unchanged, the MTTR, which describes the average failure length, is varied. The results, shown in Figure 6-15, indicate that an underestimation of the average failure length does not change the accuracy of the Digital Twin, but that in case of an overestimation of more than 10% the accuracy decreases.

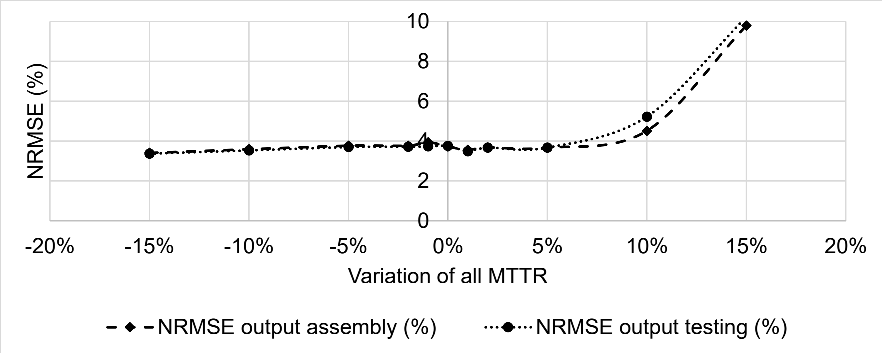


Figure 6-15 NRMSE when MTTR are varied

An explanation of the results would be that in the system under consideration shorter but therefore more failures have the same effect as less but longer failures. Thus, the system is robust until a certain average failure length, but if this average failure length is reached, its performance deteriorates rapidly. This might occur because the system reaches a tipping point where the available buffers inside the system are insufficient to absorb the negative effects of the long machine failures. Buffers are normally designed to protect the system against failures of a certain length but when this length is surpassed, the whole system becomes blocked.

6.2.1.3 Material flow

To demonstrate how the importance of the correct information about the material flow can be assessed, the accuracy of several weeks with correct material flow and with altered material flows (machines were missing in material flow) was compared. As the results in Figure 6-16 indicate, the influence of the material flow data on the accuracy of the Digital Twin can be massive. Many times a simulation model will not even be executable if the material flow or other dynamic behavior is modeled incorrectly. Of course not all possible alterations of material flow data can be analyzed or even enumerated, but Figure 6-16 should give an idea of how its importance could be examined. In this example, some stations are not included in the material flow, which can in reality happen when the material flow is deducted purely based on electronic data and some machines do not have a PLC themselves or are not connected to the MES. These stations might then be ignored by the material flow recognition.

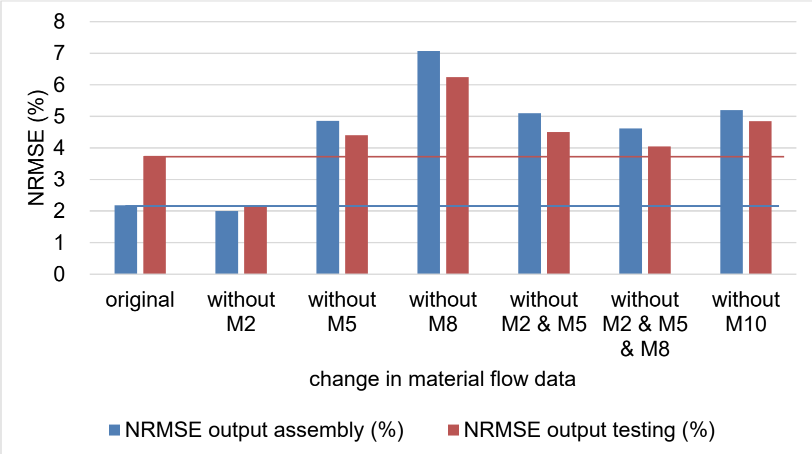


Figure 6-16 Effect of changes in material flow data on accuracy of the Digital Twin

It is striking that the accuracy becomes even better in one case if one machine is not included in the material flow. This can be caused by an overestimation of the process times in general which is in this case levelled out by including one process step less. This is an example of a bad model giving better results for a certain accuracy metric than a better model, because modeling and data errors compensate each other. This is why it is crucial to always consider multiple accuracy metrics when assessing the validity of the Digital Twin. However, in general the accuracy of the Digital Twin deteriorates when machines are not included in the material flow. This effect is not linear, so the neglect of more machines does not always decrease the accuracy of the Digital Twin further. Therefore, it is not possible to predict the exact effect of missing machines in the material flow data.

6.2.2 Importance of individual information components

In order to investigate the importance of the different input information for the Digital Twin of the production system, parts of the information were added step-by-step to the update. Figure 6-17 shows the accuracy of the Digital Twin at each stage for an exemplary week. It is obvious that each additional information increases the accuracy of the Digital Twin. In this case, the addition of the most recent process times and of the most recent scrap rate led to a bigger decrease in relative error and NRMSE than the addition of the latest availability and MTTR information.

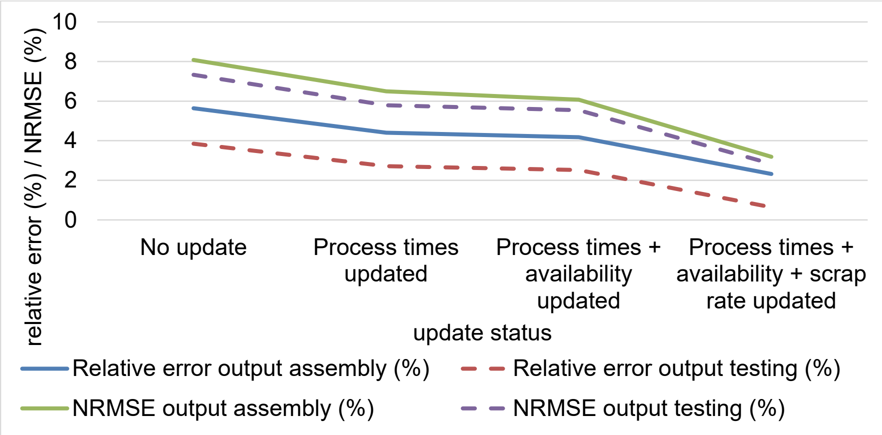


Figure 6-17 Influence of the information components on the accuracy of the Digital Twin for one exemplary week

The statements obtained by this analysis are intended to give companies whose digital data acquisition is not yet far advanced, an orientation which data are necessary for the successful implementation of Digital Twins of their production systems.

6.3 Use of the Digital Twin in practice

The Digital Twin of the production system itself does not create any monetary value for the company. Value is only generated by its use for the planning, control and operation of the production system by the responsible employees. The interactive development of a utilization concept for the Digital Twin can help to engage future users as early as possible to increase their acceptance and adoption of the new tool. A structured way to do this will be presented in the next section. In addition, examples of the successful use of the Digital Twin of the production system for production planning at Bosch will be given to present the reader insights in how a Digital Twin of the production system can be used.

6.3.1 Utilization concept

A utilization concept that describes the affected user groups, their tasks, rights and required competencies is important for the successful use of the Digital Twin. A systematic approach involving all stakeholders was developed for the Digital Twin in the industrial use case for this purpose by Overbeck et al. (2022). After documentation of possible in- and outputs of the Digital Twin and identification of the relevant stakeholders, three interactive workshops with the stakeholders are the core of the procedure.

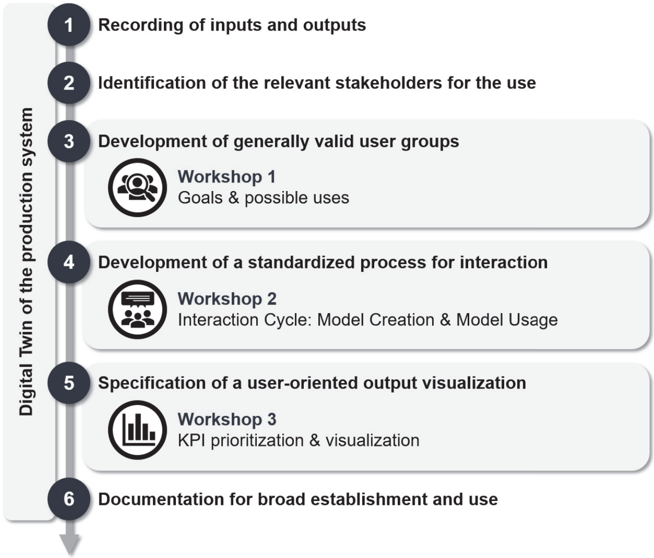


Figure 6-18 Sequence for the creation of the user concept (Overbeck et al. 2022)

As summarized in Figure 6-18, the objective of workshop 1 is to define user groups of the Digital Twin based on their objectives and possible use cases for the Digital Twin. Standardized interactions procedures are developed for each of these user groups in workshop 2. Which KPIs each user group wants to see and how they should be visualized is compiled in workshop 3.

The completed utilization concept for the Bosch use case includes an interaction cycle for model utilization including associated tasks and responsibilities which is shown in Figure 6-19, a guideline to describe the objective(s) of the Digital Twin and system, and a form to coordinate and document project-specific model requirements.

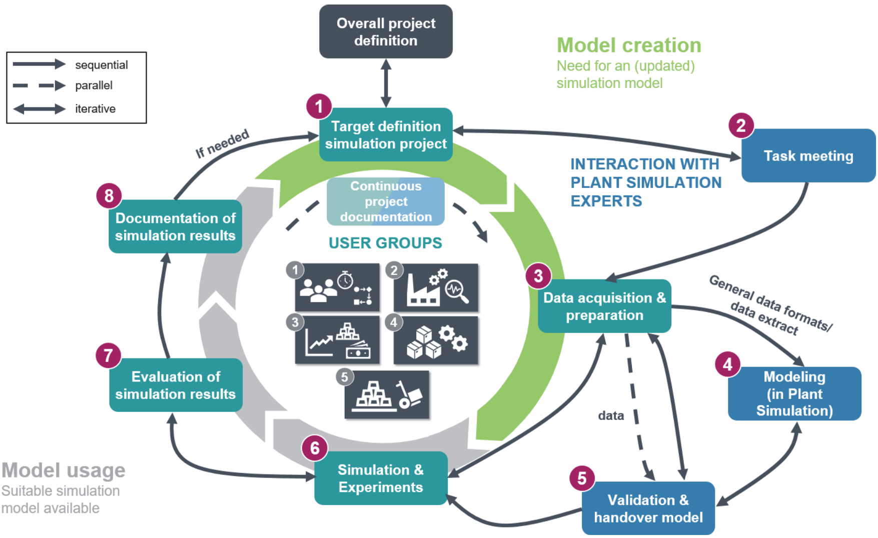


Figure 6-19 User concept Bosch use case (Overbeck et al. 2022)

The interaction cycle includes eight tasks from which tasks 2, 4, and 5 have to be performed by a simulation expert and the other tasks by the project stakeholders and users. One objective of this separation is to reduce the workload of the modeler to his core tasks and to enable the user to work with the Digital Twin on their own and establish their ownership of the Digital Twin.

6.3.2 Exemplary potentials of a Digital Twin

The implemented Digital Twin of the production system at Bosch has been used on multiple occasions over a period of several years to evaluate different scenarios and actions to either improve productivity, respond to changing external conditions, or prepare for future internal changes (e.g. introduction of new product variants). The usage of the Digital Twin in this use case is shown in Figure 6-20 following the vision of Figure 1-1. The benefits of being able to evaluate decisions in advance and avoid wrong decisions at any time are not always easy to quantify, but the following examples (which by no means claim to be exhaustive) are intended to illustrate the potential of a Digital Twin of a production system when used correctly.

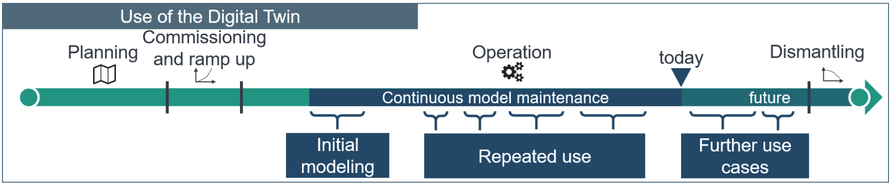


Figure 6-20 Usage of Digital Twin for analyses and improvement projects over time

That fact that it is advantageous to eliminate planning errors as early as possible in a project is a well-known fact amongst decision makers but the consequence that its early elimination makes it difficult to quantify how valuable the elimination of this error was in the end is rarely fully understood. Since simulation is often used exactly for finding and eliminating errors before they are made, its evaluation as a tool is heavily inflicted by this cognitive bias.

In the following examples for the successful usage of the Digital Twin in the industrial use case for production planning, control, and optimization are presented to demonstrate its broad applicability and its potential even in brownfield production systems.

6.3.2.1 Developing new worker instructions

As described before, the number of employees in the assembly and testing subsystems can fluctuate. In addition, depending on the product variant and the number of employees, the overall system bottleneck is either in assembly or testing. It could therefore make sense in certain situations to move one employee flexible between the two subsystems (similar to a so-called jumper), which is not yet done in the production system under consideration.

With the Digital Twin, it was possible to test various concepts for jumper deployment. A key question was how to define when the jumper should work in which subsystem. Quickly, the number of parts in the buffer between both subsystems was identified as an important indicator of when the jumper should change subsystem. But the thresholds at which the change should be performed could only be defined by simulation experiments testing different thresholds. In addition to the overall productivity of the system for different thresholds, the focus was in particular on the effects on the employee who was taking the job as jumper: How long can he work in one subsystem without having to change? Which distance does he have to walk per shift?

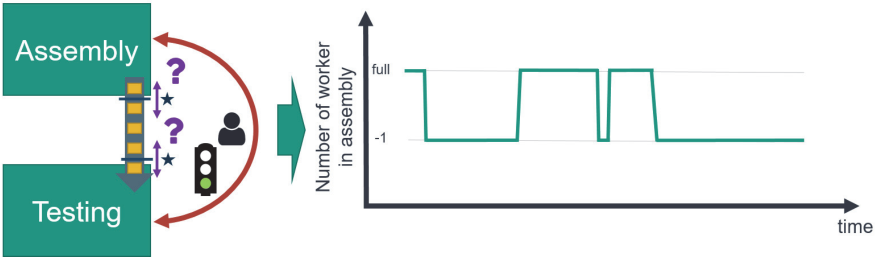


Figure 6-21 Determination of threshold to inform worker when to change subsystem

Thus, the best threshold could be identified, which showed an improvement in productivity of 21% for certain numbers of employees compared to the state without jumper. As depicted in Figure 6-21, there are certain episodes in which the jumper remains in one subsystem rather short, but in general he or she stays in each subsystem for sufficient time.

6.3.2.2 Efficient worker deployment in different demand scenarios

As some market scenarios predict a declining demand for the produced product, this could lead to an underutilization of the existing production lines in the upcoming years. The operational goal of the planning department could therefore shift from achieving the highest possible productivity and utilization of the machines to the objective of achieving the required production volume at the lowest cost possible. This can be achieved by transferring workers from the lines to other production systems of the plant to reduce the number of employees in the line. As the planned cycle times for each variant become more imprecise with a reduced number of workers on the line (because it differs further from the planned state), the planning of the required workforce for the next years becomes more imprecise and unreliable, resulting in higher safety margins and therefore in overstaffing, which causes additional costs. This can be prevented with the Digital Twin of the production system, which enables more accurate estimations of cycle times and therefore better capacity planning. With the use of the Digital Twin, ways to reduce the planned work force by 12% were found while still being able to fulfill the required production volume. The developed solution included the shifting of production orders between production lines where they can be produced most efficiently and the optimal distribution of worker between lines and subsystems. (Overbeck et al. 2023)

With the Digital Twin of the production system, it was also possible to optimize work instructions for reduced numbers of workers through simulation experiments and without expensive work by an expert, as shown in Figure 6-22 for the generic example from Figure 5-6. This can lead to improvements in cycle time, worker capacity needed, and thus lower operating costs. For example, it was possible to develop new worker loops that suggest a cycle time reduction of relevant product variants by 5-7%.

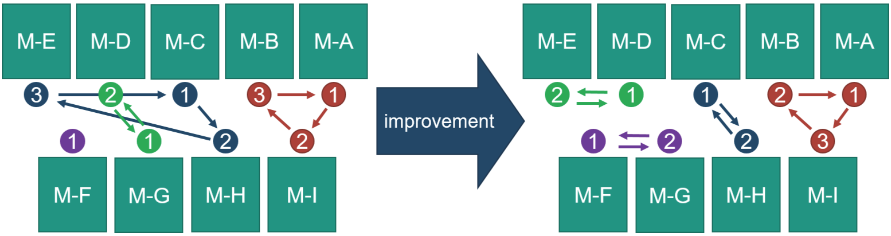


Figure 6-22 Example for improved worker loops in generic example

6.3.2.3 Targeted maintenance and machine improvement

In order to use the available maintenance staff as efficiently as possible, they can work on fundamentally eliminating machine problems which regularly trigger failures in addition to eliminating acute machine failures and performing routine maintenance. To do this, it is important to know which failures on which machines cause the greatest loss of output in the long term. This can be determined with the help of the Digital Twin of the production system. First, the machine which leads to the highest OEE, if it would run without any failures, is identified, as shown in Figure 6-23 (here machine 7). Then it can be estimated how much the productivity would increase if each failure type on the machine would be definitively eliminated (Figure 6-24). This information can then be used to prioritize which fundamental machine problems and failure types should be addressed first to increase line efficiency. In a second step, it can be evaluated which investments are economically justified to eliminate the failure causes. Therefore, the costs or lost profits of the shortfalls per year caused by this specific malfunction must be quantified in monetary terms and compared to the necessary investment for its elimination.

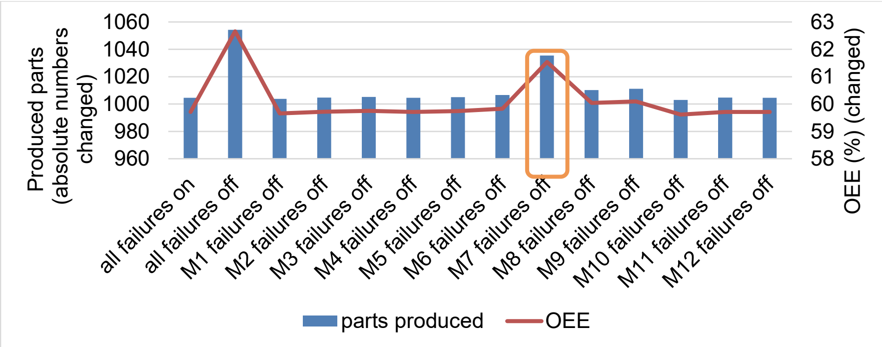


Figure 6-23 Influence of failures on different machines on the performance of the production system

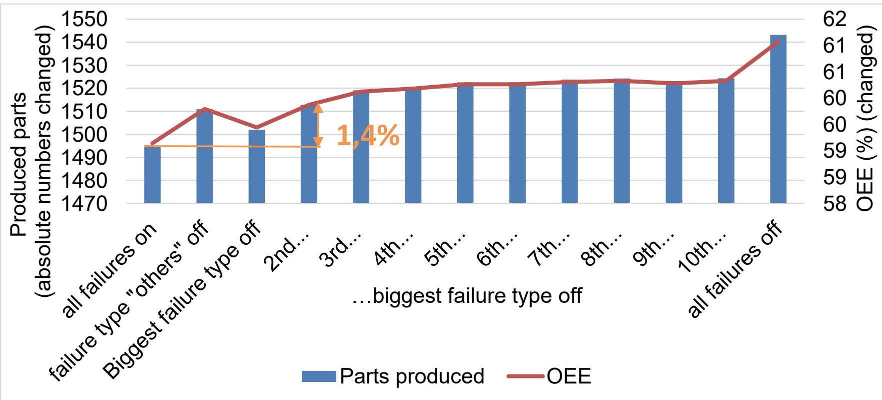


Figure 6-24 Influence of the most important failure types of machine 7 on the performance of the production system

6.3.2.4 Evaluation of hardware investments

With the help of the Digital Twin, it was possible to evaluate the effects of investing in a dedicated system for transporting and loading one certain part automatically within the line. By automating the transport of this particular part, the responsible employee could be relieved of the need to walk to this machine and could be all the time at the machine on the right, as shown in Figure 6-25. Since the machine on the right is a bottleneck machine the better availability of the employee led to an increase in the utilization of this machine, which in turn led to an increase in the OEE of the whole system by more than 4%. The decision whether it is beneficial to procure the system for automated transportation and loading of this part and to carry out the further necessary adaptions of the line to integrate it can now be made more objectively since the achievable benefit became clearly quantifiable by the Digital Twin (its costs are easy to determine also without the Digital Twin).

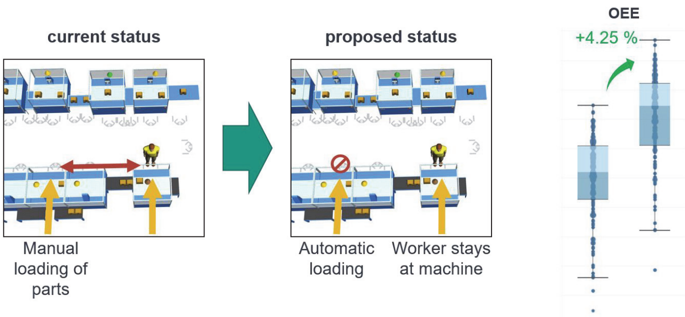


Figure 6-25 Example for the evaluation of hardware changes in the production system

6.3.2.5 Evaluation of possible short-term measures in a pandemic situation

An example of how the Digital Twin of the production system can also be used to react to completely unexpected situations as they occurred at the beginning of the Corona pandemic situation in Germany in May 2020. Even during the lockdown, possible measures for maintaining the minimum distance between employees inside the production system could be quickly developed and simulated with the Digital Twin even in home office. The visualization of the required safety distance between the workers is shown in Figure 6-26. Not only could the effectiveness of the measures in terms of ensuring the distance be tested, but the effect of the measures on the output and efficiency of the production system could also be evaluated. The measures included both structural and organizational changes. It could be shown that the preferred adaption of the worker loop and hardware configuration led to a productivity which was 4% higher than alternative measures, even if the absolute productivity decreased in comparison with the status without any pandemic measures. The measures developed and evaluated could not, of course, be implemented immediately or even automatically on site, but they do support the elaboration and objective pre-selection of ideas, providing a quantitative basis for discussion, and, therefore, enabling a faster implementation on site. In such an exceptional situation, the Digital Twin cannot replace a further review and detailed elaboration of the measures on site, but it can significantly accelerate and improve them.

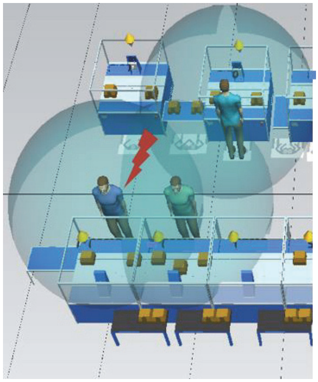


Figure 6-26 Visualization of safety distance between worker during corona pandemic

7 Discussion and outlook

The previous chapters have described the developed approach, its application to a real use case and its thorough analysis and usage. In the following chapter, the advantages and disadvantages of the proposed approach to extend common material flow simulation models into Digital Twins of the production system will be discussed.

Table 7-1 summarizes the fulfillment of the requirements that were defined based on the motivation of this research in section 1.2 in the same way they were used to evaluate the current state of research in section 3.8. When looking at data acquisition and processing it can be stated that the input of parameters (and their probabilistic distribution) can be included automatically. While dynamic behavior cannot be discovered completely automatically, two new approaches for its discovery were presented. First, it was shown how the material flow can be obtained using process mining techniques. Second, a methodology for the discovery of the responsibilities of workers for machines and their working routines using in-door localization data was introduced. As for dynamic information data, it is not yet feasible to discover and integrate structural data into simulation models. However, the presented work enhanced the possibilities and demonstrated new ways to increase the automation degree. These are based on common MES data as well as additional localization data. Because not all information concerning dynamic behavior and system structure can be obtained automatically from the available data, the two corresponding circles are only half-filled.

In contrast to the mostly separated consideration of the input data components in previous research, the presented work provides a holistic approach for all information types needed for automated simulation updating.

Table 7-1 Fulfillment of defined requirements by the presented approach

Requirement	1. data acquisition and processing			2. procedure for validation and adaptation			3. study of Digital Twin		
Aspects of requirement	1.1 Parameter	1.2 Dynamic behavior	1.3 Structure	2.1 Automated validation	2.2 Automated update	2.3 Directly applicable to real data	3.1. Validation on real use case	3.2 Accuracy over time	3.3. Influence of available data
Proposed approach	●	◑	◑	●	◕	◕	●	●	●

As shown in Table 7-1, this work includes a completely automated validation procedure, which is also capable of triggering automatically required model updates and this way closing the Digital Twin cycle. The update itself is mostly automated, but certain changes will have to be performed by a human simulation expert. Therefore, the corresponding circle is filled only to three-quarters. The algorithms include mechanisms for data cleaning and are thus applicable directly to real data from the IT systems of the company. Since for some input information additional data sources are required the circle is not filled completely. An analysis concept for the temporal behavior of the Digital Twin is developed and performed for the industrial use case. It was shown that the Digital Twin methods improve and maintain the accuracy of the simulation model. The influence of the availability and the quality of real data was evaluated in the real use case. This analysis provides insights for managers into which data and of data quality are required to be able to create meaningful Digital Twins of their production systems.

The developed approach could be successfully applied to a real-world use case that included multiple lines and therefore allowed a broad evaluation of possible challenges. After implementation, the Digital Twin could be used for various planning tasks and demonstrated its potential for more efficient and flexible production. To sum up, all analytical requirements are fully satisfied.

The presented approach fulfills the posed requirements to a large extent, provides guidance for the creation of Digital Twins of production system, and gives insights into their behavior and requirements. Yet, there remains a lot of potential for further research and improvement.

Because the progression of data collection and connectivity is heterogeneous between companies and sometimes even between different plants of one company, the approach should be applied to more production systems to supplement the presented methods for data preparation and input calculation. The proposed methodology of validation and updating can remain unchanged while the considered validation metrics and their respective limits might be adapted.

The presented way of analyzing the behavior of the Digital Twin can be further amplified and adapted to each use case. A deeper analysis of requirements of the Digital Twin is necessary for the further development of the presented approach into a commercial software solution. An intensified analysis of the Digital Twin and its dependencies on the data quantity, granularity, and quality is a precondition for the development of an off-the-shelf solution based on the research presented in this work.

Another important aspect when applying the approach is the integration of the Digital Twin of the production system into the business processes and organization of the company. To unfold the full potential of the Digital Twin of the production system, it has to be used regularly for long- and short-term planning tasks. Therefore, it has to become an integral part of regular management routines such as shop floor management.

Improvement potential lies in the use of ontologies for the description of the company data. If data are enhanced by metadata and its interconnections are systematically described and available, the transfer of the Digital Twin concept from one production system to another would be greatly facilitated.

More research efforts should be made to investigate the Digital Twin of the production system especially for companies with a low level of data acquisition and processing. For this purpose, further methods for extracting information from incomplete and/or error-prone data sets can be used. The further development of the approach for Digital Twins of Production Systems could also include the recognition of additional decision rules and dynamic behaviors in the production system.

Further potential for improving the Digital Twin of production systems arises from the continuing digitalization of companies in general and of production in particular. Better, more finely granular, and user-specific data collection can further improve the development and application of Digital Twins, as the sensitivity analysis in section 6.2 has pointed out, because better data lead to higher accuracy of the model. This applies in particular to data that are not currently collected systematically but are essential for a holistic representation of the production system, such as localization data, as well as the explicit documentation of decision rules. Further algorithms (potentially from the field of Artificial Intelligence) have to be developed to deduce these rules from data.

To extend the life span of Digital Twins even further, it would be desirable to extend the developed procedure and its methods so that they support the transition from simulation models created during the planning of a new production system into Digital Twins that accompany the operation of the built and commissioned production system. For this purpose, the procedure described would first have to be considered in some application cases already during the simulation model creation of so-called greenfield projects and it would have to be investigated which adaptations in data collection and analysis, modeling, and implementation are necessary to work with planning data.

To fulfill the vision of the virtual factory and, in the next step, the virtual company, it is necessary to integrate Digital Twins on different levels, i.e., of plants, machines, or complete production networks in a Digital Twin ecosystem.

8 Summary

After pointing out the need for a holistic approach for Digital Twins of production systems from industrial perspective as well as research perspective, the research objective to develop an approach to turn material flow simulation models into Digital Twins of production systems through automated model validation and model update using real data was formulated. To give the reader the necessary background information for the understanding of the developed approach, the required fundamentals in Digital Twins, simulation, production system modeling, and data handling in production including process mining were briefly summarized.

Next, the existing research literature on related topics was reviewed for similar approaches. Existing works were categorized in the categories: Digital Twin classifications and concepts, simulations input data management, simulation model validation and verification, automated model generation and parametrization, CMSD, detection of dynamic behavior and detection of system structure. It was shown that no existing approach satisfies all necessary requirements. In order to close the identified research gap in literature, a procedure for the enhancement of material flow simulation models into Digital Twins of production systems was developed.

After presenting the general procedure and the considerations on the initial seed model, the central functionalities for model validation and model update, including the required algorithms for data processing and input information computation were presented in detail. In particular, new approaches for the recognition of the material flow by means of process mining and the use of localization systems were proposed as well as procedures to gain insights into the structure of the production system based on production data. Subsequently, the developed concept was implemented in an industrial use case at the company Bosch as well as in the laboratory environment of the learning factory at the wbk Institute of Production Science. All modeling and configuration decisions of the industrial use case were discussed.

After implementation, the Digital Twin was intensively analyzed to understand its behavior in the industrial setting and to gain insights into the dependence of the achievable accuracy on the available data quantity and quality. It was shown that the proposed approach leads to a higher accuracy of the Digital Twin, which was analyzed with various accuracy metrics. Furthermore, the analysis indicated that the chosen update period length of one month lead to good accuracy results and that the repetition of validation (and if necessary update) every week ensures a high level of accuracy of the Digital Twin. A sensitivity analysis indicated that the performance of the Digital Twin is robust in case of small deviations in data quality of e.g. less than +5% for the mean of machine process time. It proved to be insensitive to errors in the estimation of standard deviation and the distribution type (when mean and standard deviation are correct and depending on the selection of distribution types). Any decrease in machine availability as well as an increase of MTTR of more than 10% showed both a strong effect on the accuracy of the Digital Twin. An example for the analysis of the importance of individual components of the simulation input information indicated that in the period under consideration the update of availability information had no influence on accuracy whereas new information on process times and scrap rate increased accuracy.

Subsequently, the usage of the Digital Twin during the operation of the production system was presented using a procedure for the development of a utilization concept and examples of performed experiments and investigations. The examples demonstrated the potential of Digital Twins for the definition of new worker loops (which showed productivity increases of up to 21% for certain situations), the efficient deployment of worker in different scenarios (allowing a 12% productivity increase in comparison with previous planning), and the efficient use of available maintenance capacity for machine improvement. The Digital Twin was also used to assess the potential benefits of hardware investments to quantify investment decision in an example which showed a potential efficiency improvement of more than 4%. The possibility to use the Digital Twin to react quickly in unforeseen situations was demonstrated during the corona pandemic during which different scenarios for the implementation of personal protection measures could be evaluated rapidly.

The presented approach answers the research requirements, which arose from the motivation and the state of research. After discussing the benefits of the approach for Digital Twins of production systems, further research possibilities are identified which include its application to more use case in order to expand the tool box of methods for data retrieval, processing, and simulation input computation in new environments. This might also include the consideration of other accuracy metrics and limit values. Additionally, the integration of the Digital Twin of the production system with other Digital Twins and into the business processes of the company itself would leverage its potential on the journey to the truly digitalized company.

Own publications

Brützel, O.; Küppers, F.; Overbeck, L.; Stricker, N.; Verhaelen, B. & Lanza, G. (2021), „Eine automatisierungsgerechte robuste Produktionsplanung : Robustheit der Planung im Netzwerk unter Auftragsunsicherheit“, *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 116(1-2), pp. 44–48. https://doi.org/10.1515/zwf-2021-0009.

Brützel, O.; Overbeck, L.; Nagel, M.; Stricker, N. & Gisela, L. (2020), „Generische Modellierung von halbautomatisierten Produktionssystemen für Ablaufsimulationen“, *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 115(11), pp. 792–796.

Brützel, O.; Voelkle, D.; Overbeck, L.; Stricker, N. & Lanza, G. (2022), „Automated Production Network Planning Under Uncertainty by Developing Representative Demand Scenarios“. *Towards Sustainable Customization: Bridging Smart Products and Manufacturing Systems*, eds. A.-L. Andersen, R. Andersen, T. D. Brunoe, M. S. S. Larsen, K. Nielsen, A. Napoleone & S. Kjeldgaard, Springer International Publishing, Cham, pp. 459–466. ISBN: 978-3-030-90700-6.

Brützel, O.; González Di Miele, Román Ignacio; Overbeck, L.; May, M. C. & Lanza, G. (2022), „Modeling of a Modular Discrete Event Simulation for Fuel Cell Assembly within a Factory Model“ in FC3 Fuel Cell Conference Chemnitz, Chemnitz.

Fleischer, J.; Lanza, G.; Wirth, F.; Gönnheimer, P.; Peukert, S.; May, M.; Hausmann, L.; Fraider, F.; Netzer, M.; Oexle, F.; Silbernagel, R. & Overbeck, L. (2021), Leitfaden Antriebstechnik 4.0: Digitalisierungstrends für Produkt, Produktion und Lieferkette, Frankfurt.

Greinacher, S.; Overbeck, L.; Kuhnle, A.; Krahe, C. & Lanza, G. (2020), „Multi-objective optimization of lean and resource efficient manufacturing systems“, *Production Engineering*, 14(2), pp. 165–176. https://doi.org/10.1007/s11740-019-00945-9.

Kimmig, A.; Schöck, M.; Mühlbeier, E.; Oexle, F.; Fleischer, J.; Bönsch, J.; Ovtcharova, J.; Hahn, J.; Grunwald, A.; Albers, A.; Rapp, S.; Hagenmeyer, V.; Scholz, S. G.; Schmidt, A.; Müller, T.; Becker, J.; Schade, F.; Beyerer, J.; Rehak, J.; Zwick, T.; Pauli, M.; Nuß, B.; Lanza, G.; Schild, L. & Overbeck, L. (2021), „Wertstromkinematik – Produktionssysteme neu gedacht: Interdisziplinäres Forscherteam arbeitet an der Produktionstechnik der Zukunft (Teil 2)“, *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 116(12), pp. 935–939. https://doi.org/10.1515/zwf-2021-0207.

Kirchberger, M.; Heeger, M.; Altay, A.; Liebrecht, C.; Overbeck, L.; Kandler, M.; Lanza, G.; Voigt, C. & Franke, J. (2022), „Simulationsgestütztes Vorgehensmodell zur Realisierung einer Matrixfertigung: Planung, Bewertung und Digitaler Zwilling in der Elektronikindustrie bei der Siemens AG Karlsruhe“, *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 117(4), pp. 224–228. https://doi.org/10.1515/zwf-2022-1037.

May, M. C.; Overbeck, L.; Wurster, M.; Kuhnle, A. & Lanza, G. (2021), „Foresighted digital twin for situational agent selection in production control“, *Procedia CIRP*, 99, pp. 27-32. https://doi.org/10.1016/j.procir.2021.03.005.

Overbeck, L.; Brützel, O.; Teufel, M.; Stricker, S.; Kuhnle, A. & Lanza, G. (2021), „Continuous adaption through real data analysis turn simulation models into digital twins“, *Procedia CIRP*, 104, pp. 98–103. https://doi.org/10.1016/j.procir.2021.11.017.

Overbeck, L.; Brützel, O.; Stricker, S. & Lanza, G. (2020), „Digitaler Zwilling des Produktionssystems*“, ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 115(Special), pp. 62–65. https://doi.org/10.3139/104.112326.

Overbeck, L.; Le Louarn, A.; Brützel, O.; Stricker, N. & Lanza, G. (2022), „Comprehensive Validation Metrics and Precise Updating of Digital Twins of Production Systems“, *SNE Simulation Notes Europe*, 32(3), pp. 135–142. https://doi.org/10.11128/sne.32.tn.10613.

Overbeck, L.; Baralija, D.; Silbernagel, R.; Rauen, H.; Fleischer, J. & Lanza, G. (2022), „Digitalisierung in der Supply Chain: Trends und Anwendungsfelder in der Branche der Antriebstechnik des Maschinen- und Anlagenbaus“, *Industrie 4.0 Management*, 2022(4), pp. 23–27.

Overbeck, L.; Rose, A.; May, M. & Lanza, G. (2022), „Nutzungskonzept für Digitale Zwillinge von Produktionssystemen“, *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 117(4), pp. 244–248. https://doi.org/10.1515/zwf-2022-1035.

Overbeck, L.; Hugues, A.; May, M. C.; Kuhnle, A. & Lanza, G. (2021), „Reinforcement Learning Based Production Control of Semi-automated Manufacturing Systems“, *Procedia CIRP*, 103(7676), pp. 170–175. https://doi.org/10.1016/j.procir.2021.10.027.

Overbeck, L.; Le Louarn, A.; Brützel, O.; Stricker, N. & Lanza, G. (2021), „Continuous Validation and Updating for High Accuracy of Digital Twins of Production Systems" in *Simulation in Produktion und Logistik 2021*, eds. J. Franke & P. Schuderer, Cuvillier Verlag, Göttingen, pp. 609–617.

Overbeck, L.; Graves, S. C. & Lanza, G. (2023), „Development and analysis of digital twins of production systems", *International Journal of Production Research.* https://doi.org/10.1080/00207543.2023.2242525

Overbeck, L.; Glaser, V.; May, M. C. & Lanza, G. (2023), „Generalization of Reinforcement Learning Agents for Production Control" in *Proceedings of the Changeable, Agile, Reconfigurable and Virtual Production Conference and the World Mass Customization & Personalization Conference*. Springer, Cham, pp. 338-346. http://dx.doi.org/10.1007/978-3-031-34821-1_37

Overbeck, L.; Voigtländer, S. & Lanza, G. (2023), „ Optimal line configurations for agile production systems for battery cell manufacturing " in *Procedia CIRP, 118,* pp. 976-981. https://doi.org/10.1016/j.procir.2023.06.168

Ruhland, J.; Storz, T.; Kößler, F.; Ebel, A.; Sawodny, J.; Hillenbrand, J.; Gönnheimer, P.; Overbeck, L.; Lanza, G.; Hagen, M.; Tübke, J.; Gandert, J.; Paarmann, S.; Wetzel, T.; Mohacsi, J.; Altvater, A.; Spiegel, S.; Klemens, J.; Scharfer, P.; Schabel, W.; Nowoseltschenko, K.; Müller-Welt, P.; Bause, K.; Albers, A.; Schall, D.; Grün, T.; Hiller, M.; Schmidt, A.; Weber, A.; Biasi, L. de; Ehrenberg, H. & Fleischer, J. (2022), „Development of a Parallel Product-Production Co-design for an Agile Battery Cell Production System". *Towards Sustainable Customization: Bridging Smart Products and Manufacturing Systems*, eds. A.-L. Andersen, R. AndersenBrunoe, T. Ditlev, S. Larsen, M. Stoettrup, K. Nielsen, A. Napoleone & S. Kjeldgaard, Springer, Cham, pp. 96–104. https://doi.org/10.1007/978-3-030-90700-6_10.

.

References

References according to the scheme (A_<last name> <year>) refer to Bachelor and Master Theses at the wbk Institute of Production Science, which were supervised by the author of this dissertation.

A_Brützel, O. 2019

Brützel, O. (2019), *Baukastensystem zur Implementierung von Simulationen automatisierter Montagesysteme nach dem Chaku-Chaku-Prinzip*. Master Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

A_Haizmann, T. 2020

Haizmann, T. (2020), *Erarbeitung eines Standards zur intelligenten, automatisierten Erstellung ereignisdiskreter Simulationsmodelle von bereits existierenden Produktionsanlagen mittels Produktionsdaten*. Master Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

A_Janikovits, M. 2020

Janikovits, M. (2020), *KI-basierte Datenanalyse zur Effizienzsteigerung in halbautomatisierten Montagezellen*. Master Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

A_Kudlik, L. 2022

Kudlik, L. (2022), *Erkennung von Maschinenpositionen anhand von Mitarbeiterlokalisierungsdaten und anschließende Produktionsoptimierung*. Bachelor Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

A_Le Louarn, A. 2021

Le Louarn, A. (2021), *Development of a concept for optimal updating of a digital twin of a production system*. Master Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

A_Liu, S. 2019

Liu, S. (2019), *Modulare Simulation einer automatisierten Montagelinie nach dem Chaku-Chaku Prinzip*. Master Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

A_Maul, N. 2021

Maul, N. (2021), *Erstellung einer Bibliothek zur Simulation von halbautomatisierten Produktionssystemen und Optimierung der Mitarbeiter- und Auftragszuteilung mit Hilfe der Simulation*. Master Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

A_Merker, C. 2020

Merker, C. (2020), *Generische Erstellung und automatisierte Pflege von ereignisdiskreten Ablaufsimulationsmodellen durch Abbildung des Systemaufbaus aus Realdaten*. Master Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

A_Nagel, M. 2020

Nagel, M. (2020), *Übertragbarkeit ereignisdiskreter Simulationsmodelle mit Hilfe generischer Modelle und Simulationsbausteinen zur Analyse paralleler Produktionslinien*. Master Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

A_Teufel, M. 2020

Teufel, M. (2020), *Datenbasierte Ableitung von Ablauflogiken in Produktionssystemen für die Adaption von Simulationsmodellen*. Master Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

A_Xie, X. 2020

Xie, X. (2020), *Entwicklung und Evaluation der Prognosefähigkeit einer Ablaufsimulation für Produktionssysteme*. Master Thesis, Karlsruher Institut für Technologie (KIT), Karlsruhe, wbk Institut für Produktionstechnik.

Acél 1992

Acél, P. (1992), „Systems-Engineering und Simulation“ in *Simulation. Ein Blick in die Zukunft,* eds. BW/FOR der ETH Zürich.

Albrecht et al. 2014

Albrecht, F.; Kleine, O. & Abele, E. (2014), „Planning and Optimization of Changeable Production Systems by Applying an Integrated System Dynamic and Discrete Event Simulation Approach“, *Procedia CIRP*, 17, pp. 386–391. https://doi.org/10.1016/j.procir.2014.01.039.

Analla 1998

Analla, M. (1998), „Model validation through the linear regression fit to actual versus predicted values“, *Agricultural Systems*, 57(1), pp. 115–119. https://doi.org/10.1016/s0308-521x(97)00073-5.

Anderson & Darling 1954

Anderson, T. W. & Darling, D. A. (1954), „A Test of Goodness of Fit“, *Journal of the American Statistical Association*, 49(268), pp. 765–769. https://doi.org/10.2307/2281537.

Andreasson et al. 2019 - 2019

Andreasson, H.; Weman, J.; Nafors, D.; Berglund, J.; Johansson, B.; Lihnell, K. & Lydhig, T. (2019 - 2019), „Utilizing Discrete Event Simulation to Support Conceptual Development of Production Systems“. *Proceedings of the Winter Simulation Conference,* eds. N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas & Y.-J. Son, IEEE, pp. 2073–2084. ISBN: 978-1-7281-3283-9. https://doi.org/10.1109/wsc40007.2019.9004943.

Angles et al. 2018

Angles, R.; Arenas, M.; Barceló, P.; Hogan, A.; Reutter, J. & Vrgoč, D. (2018), „Foundations of Modern Query Languages for Graph Databases“, *ACM Computing Surveys*, 50(5), pp. 1–40. https://doi.org/10.1145/3104031.

Aufenanger et al. 2010

Aufenanger, M.; Blecken, A. & Laroque, C. (2010), „Design and implementation of an MDA interface for flexible data capturing“, *Journal of Simulation*, 4(4), pp. 232–241. https://doi.org/10.1057/jos.2010.13.

Autiosalo et al. 2020

Autiosalo, J.; Vepsalainen, J.; Viitala, R. & Tammi, K. (2020), „A Feature-Based Framework for Structuring Industrial Digital Twins“, *IEEE Access*, 8, pp. 1193–1208. https://doi.org/10.1109/ACCESS.2019.2950507.

Aydt et al. 2008

Aydt, H.; Turner, S. J.; Cai, W. & Low, M. Y. H. (2008), „Symbiotic Simulation Systems: An Extended Definition Motivated by Symbiosis in Biology“. *22nd Workshop on Principles of Advanced and Distributed Simulation,* pp. 109–116. ISBN: 978-0-7695-3159-5. https://doi.org/10.1109/PADS.2008.17.

Balci 1989

Balci, O. (1989), „How to assess the acceptability and credibility of simulation results". *1989 Winter Simulation Conference proceedings,* eds. E. A. MacNair, K. J. Musselman & P. Heidelberger, pp. 62–71. https://doi.org/10.1145/76738.76746.

Balci 1998

Balci, O. (1998), „Verification, Validation, and Testing" in *Handbook of Simulation,* eds. J. Banks, John Wiley & Sons, Inc, Hoboken, NJ, USA, pp. 335–393.

Bamberg 2017

Bamberg, G. (2017), *Statistik*. *Eine Einführung für Wirtschafts- und Sozialwissenschaftler,* De Gruyter, Berlin/Boston. ISBN: 9783110495720.

Banks et al. 2010

Banks, J.; Carson, J. S.; Nelson, B. L. & Nicol, D. M. (2010), *Discrete-event system simulation,* Pearson Education, Upper Saddle River, N.J., London. ISBN: 9780138150372.

Bao et al. 2018

Bao, J.; Guo, D.; Li, J. & Zhang, J. (2018), „The modelling and operations for the digital twin in the context of manufacturing", *Enterprise Information Systems*, 13(4), pp. 534–556. https://doi.org/10.1080/17517575.2018.1526324.

Barlas 2015

Barlas, P. (2015), *Automating Simulation: an Open Source Software for Automated Input Data in Discrete Event Simulation Projects*. Dissertation, University of Limerick, Limerik.

Barricelli et al. 2019

Barricelli, B. R.; Casiraghi, E. & Fogli, D. (2019), „A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications", *IEEE Access*, 7, pp. 167653–167671. https://doi.org/10.1109/ACCESS.2019.2953499.

Barring et al. 2018

Barring, M.; Johansson, B.; Flores-Garcia, E.; Bruch, J. & Wahlstrom, M. (2018), „Challenges of data acquisition for simulation models of production systems in need of standards". *2018 Winter Simulation Conference (WSC),* eds. M. Rabe, N. Mustafee, A. Skoogh, S. Jain & B. Johansson, IEEE, pp. 691–702. https://doi.org/10.1109/wsc.2018.8632463.

Barth et al. 2009

Barth, M.; Strube, M.; Fay, A.; Weber, P. & Greifeneder, J. (2009), „Object-oriented engineering data exchange as a base for automatic generation of simulation models“. *2009 35th Annual Conference of IEEE Industrial Electronics,* eds. A. Carvalho, C. Cuoto, L. Gomes & J. J. Rodriguez Andina, IEEE, pp. 2465–2470. https://doi.org/10.1109/iecon.2009.5415229.

Becker et al. 1995

Becker, J.; Rosemann, M. & Schütte, R. (1995), „Grundsätze ordnungsmäßiger Referenzmodellierung“, *Wirtschaftsinformatik*, 37(5), pp. 37–110. https://doi.org/10.1007/978-3-663-10233-5.

Becker 2018

Becker, T. (2018), *Prozesse in Produktion und Supply Chain optimieren,* Springer Vieweg, Berlin, Heidelberg.

Bergmann et al. 2017

Bergmann, S.; Feldkamp, N. & Straßburger, S. (2017), „Emulation of control strategies through machine learning in manufacturing simulations“, *Journal of Simulation*, 11(1), pp. 38–50. https://doi.org/10.1057/s41273-016-0006-0.

Bergmann et al. 2015

Bergmann, S.; Feldkamp, N. & Straßburger, S. (2015), „Approximation of dispatching rules for manufacturing simulation using data mining methods“ in *2015 Winter Simulation Conference WSC,* pp. 2329–2340.

Bergmann & Straßburger 2020

Bergmann, S. & Straßburger, S. (2020), „Automatische Modellgenerierung – Stand, Klassifizierung und ein Anwendungsbeispiel“ in *Ablaufsimulation in der Automobilindustrie,* eds. G. Mayer, C. Pöge, S. Spieckermann & S. Wenzel, Springer, Berlin, Heidelberg, pp. 333–347.

Bergmann 2013

Bergmann, S. (2013), *Automatische Generierung adaptiver Modelle zur Simulation von Produktionssystemen*. Dissertation, TU Ilmenau, Ilmenau.

Bianconi et al. 2020

Bianconi, C.; Bonci, A.; Monteriu, A.; Pirani, M.; Prist, M. & Taccari, L. (2020),

„System Thinking Approach for Digital Twin Analysis“. *IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC),* eds. Y. Rezgui, IEEE, Piscataway, NJ. ISBN: 9781728134017.

Bichler et al. 2017

Bichler, K.; Krohn, R.; Philippi, P. & Schneidereit, F. (2017), *Kompakt-Lexikon Logistik*. *2.250 Begriffe nachschlagen, verstehen, anwenden,* Springer Gabler Verlag, Wiesbaden. ISBN: 9783658125356.

Bokrantz et al. 2018

Bokrantz, J.; Skoogh, A.; Lämkull, D.; Hanna, A. & Perera, T. (2018), „Data quality problems in discrete event simulation of manufacturing operations“, *Simulation*, 94(11), pp. 1009–1025. https://doi.org/10.1177/0037549717742954.

Boukerche et al. 2020

Boukerche, A.; Zheng, L. & Alfandi, O. (2020), „Outlier Detection: Methods, models, and classification“, *ACM Computing Surveys*, 53(3), pp. 1–37. https://doi.org/10.1145/3381028.

Bracht et al. 2018

Bracht, U.; Geckler, D. & Wenzel, S. (2018), *Digitale Fabrik*. *Methoden und Praxisbeispiele,* Springer Vieweg, Berlin. ISBN: 978-3-662-55782-2.

Bruckner et al. 2020

Bruckner, L.; Oppelt, M.; Urbas, L. & Barth, M. (September 2020), *The current and future use of simulation in discrete and process industries*. *Results of a global online survey*, Siemens AG, Munich

Brützel et al. 2020

Brützel, O.; Overbeck, L.; Nagel, M.; Stricker, N. & Gisela, L. (2020), „Generische Modellierung von halbautomatisierten Produktionssystemen für Ablaufsimulationen“, *ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 115(11), pp. 792–796.

Carson 2002

Carson, J. S. (2002), „Model verification and validation“. *Proceedings of the Winter Simulation Conference,* eds. E. Yücesan, C.-H. Chen, J. L. Snowdon & J. M. Charnes, IEEE, pp. 52–58. ISBN: 0-7803-7614-5. https://doi.org/10.1109/WSC.2002.1172868.

Carson 1989

Carson, J. S. (1989), „Verification and validation: a consultant's perspective". *Proceedings of the 21st conference on Winter simulation,* eds. E. A. MacNair, K. J. Musselman & P. Heidelberger, ACM Press, New York, New York, USA. https://doi.org/10.1145/76738.76810.

Cheng et al. 2018

Cheng, Y.; Chen, K.; Sun, H.; Zhang, Y. & Tao, F. (2018), „Data and knowledge mining with big data towards smart production", *Journal of Industrial Information Integration*, 9, pp. 1–13. https://doi.org/10.1016/j.jii.2017.08.001.

Choi et al. 2022

Choi, T.-M.; Kumar, S.; Yue, X. & Chan, H.-L. (2022), „Disruptive Technologies and Operations Management in the Industry 4.0 Era and Beyond", *Production and Operations Management*, 31(1), pp. 9–31. https://doi.org/10.1111/poms.13622.

Cimino et al. 2019

Cimino, C.; Negri, E. & Fumagalli, L. (2019), „Review of digital twin applications in manufacturing", *Computers in Industry*, 113, pp. 103130. https://doi.org/10.1016/j.compind.2019.103130.

Collisi 2002

Collisi, T. (2002), *Ein informationslogistisches Architekturkonzept zur Akquisition simulationsrelevanter Daten,* Meisenbach, Bamberg. ISBN: 3875251644.

Damjanovic-Behrendt & Behrendt 2019

Damjanovic-Behrendt, V. & Behrendt, W. (2019), „An open source approach to the design and implementation of Digital Twins for Smart Manufacturing", *International Journal of Computer Integrated Manufacturing*, 32(4-5), pp. 366–384. https://doi.org/10.1080/0951192X.2019.1599436.

Dias et al. 2016

Dias, L. M. S.; Vieira, A. A. C.; Pereira, G. A. B. & Oliveira, J. A. (2016), „Discrete simulation software ranking - A top list of the worldwide most popular and used tools". *2016 Winter Simulation Conference (WSC),* eds. T.M.K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka & S. E. Chick, IEEE, Piscataway, NJ. https://doi.org/10.1109/wsc.2016.7822165.

DIN IEC 2014

DIN IEC (2014), *Internationales Elektrotechnisches Wörterbuch - Teil 351: Leittechnik,* Beuth, Berlin. DIN IEC 60050-351.

Ding et al. 2019

Ding, K.; Chan, F. T.S.; Zhang, X.; Zhou, G. & Zhang, F. (2019), „Defining a Digital Twin-based Cyber-Physical Production System for autonomous manufacturing in smart shop floors“, *International Journal of Production Research*, 57(20), pp. 6315–6334. https://doi.org/10.1080/00207543.2019.1566661.

Echsler Minguillon & Lanza 2019

Echsler Minguillon, F. & Lanza, G. (2019), „Coupling of centralized and decentralized scheduling for robust production in agile production systems“, *Procedia CIRP*, 79, pp. 385–390. https://doi.org/10.1016/j.procir.2019.02.099.

Eckardt 2001

Eckardt, F. (2001), *Ein Beitrag zu Theorie und Praxis datengetriebener Modellgeneratoren zur Simulation von Produktionssystemen*. Dissertation, TU Ilmenau, Ilmenau.

Eckstein 2014

Eckstein, P. P. (2014), *Repetitorium Statistik*. *Deskriptive Statistik - Stochastik - Induktive Statistik,* Springer Gabler Verlag, Wiesbaden. ISBN: 9783658057480.

Eley 2012

Eley, M. (2012), *Simulation in der Logistik*. *Einführung in die Erstellung ereignisdiskreter Modelle unter Verwendung des Werkzeuges “Plant Simulation”,* Springer Gabler Verlag, Berlin, Heidelberg. ISBN: 9783642273728.

Ester et al. 1996

Ester, M.; Kriegel, h.-P.; Sander, J. & Xu, X. (1996), „A density-based algorithm for discovering clusters in large spatial databases with noise“. *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining,* eds. U. Fayyad, E. Simoudis & J. Han, pp. 226–231.

Fang 2015

Fang, H. (2015), „Managing data lakes in big data era: What’s a data lake and why has it became popular in data management ecosystem“. *2015 IEEE International*

Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), IEEE, pp. 820–824. https://doi.org/10.1109/cyber.2015.7288049.

Fischer & Hofer 2011
Fischer, P. & Hofer, P. (2011), *Lexikon der Informatik,* Springer Berlin Heidelberg, Berlin, Heidelberg. http://dx.doi.org/10.1007/978-3-642-15126-2. ISBN: 9783642151255.

Fisel et al. 2019a
Fisel, J.; Exner, Y.; Stricker, N. & Lanza, G. (2019), „Changeability and flexibility of assembly line balancing as a multi-objective optimization problem“, *Journal of Manufacturing Systems*, 53, pp. 150–158. https://doi.org/10.1016/j.jmsy.2019.09.012.

Fisel et al. 2019b
Fisel, J.; Duffie, N.; Moser, E. & Lanza, G. (2019), „Changeability - a frequency perspective“, *Procedia CIRP*, 79, pp. 186–191. https://doi.org/10.1016/j.procir.2019.02.043.

Flores-Garcia et al. 2018
Flores-Garcia, E.; Wiktorsson, M.; Bruch, J. & Jackson, M. (2018), „Revisiting Challenges in Using Discrete Event Simulation in Early Stages of Production System Design“. *IFIP International Conference on Advances in Production Management Systems,* Springer, Cham, pp. 534–540. https://doi.org/10.1007/978-3-319-99704-9_65.

Fournier 2011
Fournier, J. (2011), „Model building with Core Manufacturing Simulation Data“. *Proceedings of the 2011 Winter Simulation Conference (WSC),* eds. S. Jain, R. R. Creasy, J. Himmelspach, K. P. White & M. Fu, IEEE. https://doi.org/10.1109/wsc.2011.6147933.

Fritz 2007
Fritz, J. U. (2007), *Zielführende Modellierung und Analyse schlanker Fertigungssysteme mit der Digitalen Fabrik*. Dissertation, Universität des Saarlandes, Saarbrücken, Lehrstuhl für Fertigungstechnik/CAM.

Fuller et al. 2020

Fuller, A.; Fan, Z.; Day, C. & Barlow, C. (2020), „Digital Twin: Enabling Technologies, Challenges and Open Research“, *IEEE Access*, 8, pp. 108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358.

Ghosh & Sampalli 2019

Ghosh, S. & Sampalli, S. (2019), „A Survey of Security in SCADA Networks: Current Issues and Future Challenges“, *IEEE Access*, 7, pp. 135812–135831. https://doi.org/10.1109/access.2019.2926441.

Goodall et al. 2019

Goodall, P.; Sharpe, R. & West, A. (2019), „A Data-Driven Simulation to Support Remanufacturing Operations“, *Computers in Industry*(105), pp. 48–60. https://doi.org/10.1016/j.compind.2018.11.001.

Greasley & Owen 2018

Greasley, A. & Owen, C. (2018), „Modelling people’s behaviour using discrete-event simulation: a review“, *International Journal of Operations & Production Management*, 38(5), pp. 1228–1244. https://doi.org/10.1108/IJOPM-10-2016-0604.

Greenwood 2020

Greenwood, A. G. (2020), „A Specification for Effective Simulation Project Management“. *2020 Winter Simulation Conference (WSC),* eds. K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder & R. Thiesing, IEEE, pp. 2482–2492. https://doi.org/10.1109/wsc48552.2020.9383982.

Greinacher et al. 2020

Greinacher, S.; Overbeck, L.; Kuhnle, A.; Krahe, C. & Lanza, G. (2020), „Multi-objective optimization of lean and resource efficient manufacturing systems“, *Production Engineering*, 14(2), pp. 165–176. https://doi.org/10.1007/s11740-019-00945-9.

Grieves 2014

Grieves, M. W. (2014), *Digital Twin: Manufacturing Excellence through Virtual Factory Replication*. *A Whitepaper*. (no location)

Grieves & Vickers 2017

Grieves, M. & Vickers, J. (2017), „Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems“ in *Transdisciplinary Perspectives on*

Complex Systems, eds. F.-J. Kahlen, S. Flumerfelt & A. Alves, Springer International Publishing, Cham, pp. 85–113.

Günther et al. 2019

Günther, L. C.; Colangelo, E.; Wiendahl, H.-H. & Bauer, C. (2019), „Data quality assessment for improved decision-making: a methodology for small and medium-sized enterprises“, *Procedia Manufacturing*, 29, pp. 583–591. https://doi.org/10.1016/j.promfg.2019.02.114.

Gutenschwager et al. 2017

Gutenschwager, K.; Rabe, M.; Spiekermann, S. & Wenzel, S. (2017), *Simulation in Produktion und Logistik*. *Grundlagen und Anwendungen,* Springer Vieweg, Berlin, Heidelberg. ISBN: 978-3-662-55744-0.

Harper et al. 2021

Harper, A.; Mustafee, N. & Yearworth, M. (2021), „Facets of trust in simulation studies“, *European Journal of Operational Research*, 289(1), pp. 197–213. https://doi.org/10.1016/j.ejor.2020.06.043.

Hedtstück 2013

Hedtstück, U. (2013), „Ereignisorientierte Simulation diskreter Prozesse“ in *Simulation diskreter Prozesse,* Springer Vieweg, Berlin, Heidelberg, pp. 75–116.

Heitmann 1999

Heitmann, K. (1999), *Sichere Prognosen für die Produktionsoptimierung mittels stochastischer Modelle*. Dissertation, TU München, München.

Herbert et al. 2021

Herbert, M.; Selmaier, A.; Mühlmann, F.; Franke, J. & Fürst, J. (2021), „Application of discrete-event simulation for factory planning – A case study“. *Simulation in Produktion und Logistik 2021,* eds. J. Franke & P. Schuderer, Cuvillier Verlag, Göttingen, pp. 133–142.

Hildebrand et al. 2018

Hildebrand, K.; Gebauer, M.; Hinrichs, H. & Mielke, M. (eds.) (2018), *Daten- und Informationsqualität*. *Auf dem Weg zur Information Excellence,* Springer Vieweg, Wiesbaden. ISBN: 978-3-658-21994-9.

Holling & Gediga 2015

Holling, H. & Gediga, G. (2015), *Statistik – Testverfahren,* Hogrefe Verlag GmbH & Company KG, Göttingen. ISBN: 9783840923029.

Horn et al. 2005

Horn, S.; Weigert, G. & Werner, S. (2005), „Data coupling strategies in production environments". *28th International Spring Seminar on Electronics Technology: Meeting the Challenges of Electronics Technology Progress, 2005,* eds. unknown, IEEE, pp. 278–282. ISBN: 0-7803-9325-2. https://doi.org/10.1109/ISSE.2005.1491040.

Hrdliczka et al. 1997

Hrdliczka, V.; Jakobi, H.; Schumacher, R. & Wenzel, S. (1997), *Leitfaden für Simulationsbenutzer in Produktion und Logistik*. *Mitteilungen aus den Fachgruppen*.

ISO/IEC 2023

ISO/IEC (2023), *Digital Twin - Concepts and terminology*. ISO/IEC AWI 30173, ISO/IEC JTC 1/SC 41.

Jensen 2007

Jensen, S. (2007), *Eine Methodik zur teilautomatisierten Generierung von Simulationsmodellen aus Produktionsdatensystemen am Beispiel einer Job-shop-Fertigung*. Dissertation, Universität Kassel, Kassel.

Jeon & Kim 2016

Jeon, S. M. & Kim, G. (2016), „A survey of simulation modeling techniques in production planning and control (PPC)", *Production Planning & Control*, 27(5), pp. 360–377. https://doi.org/10.1080/09537287.2015.1128010.

Jones et al. 2020

Jones, D.; Snider, C.; Nassehi, A.; Yon, J. & Hicks, B. (2020), „Characterising the Digital Twin: A systematic literature review", *CIRP Journal of Manufacturing Science and Technology*, 29, pp. 36–52. https://doi.org/10.1016/j.cirpj.2020.02.002.

Kellner et al. 2020

Kellner, F.; Lienland, B. & Lukesch, M. (2020), *Produktionswirtschaft,* Springer, Berlin, Heidelberg. ISBN: 978-3-662-61445-7.

Khan et al. 2014

Khan, K.; Rehman, S. U.; Asghar, S.; Fong, S. & Sarasvady, S. (2014), „DBSCAN:

Past, present and future“. *Applications of Digital Information and Web Technologies (ICADIWT), 2014 Fifth International Conference on the,* IEEE, pp. 232–238. ISBN: 978-1-4799-2259-8. https://doi.org/10.1109/ICADIWT.2014.6814687.

Khine & Wang 2018
Khine, P. P. & Wang, Z. S. (2018), „Data lake: a new ideology in big data era“, *ITM Web of Conferences*, 17(03025). https://doi.org/10.1051/itmconf/20181703025.

Kirchhof 2016
Kirchhof, P. (2016), „Automatically generating flow shop simulation models from SAP data“. *2016 Winter Simulation Conference (WSC),* eds. T.M.K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka & S. E. Chick, IEEE, Piscataway, NJ, pp. 3588–3589.

Kleijnen 1999
Kleijnen, J. P. (1999), „Validation of models: statistical techniques and data availability“, *Proceedings of the 1999 Winter Simulation Conference*(1), pp. 647–654.

Kletti & Deisenroth 2021
Kletti, J. & Deisenroth, R. (2021), *Lehrbuch für digitales Fertigungsmanagement,* Springer, Berlin, Heidelberg. http://dx.doi.org/10.1007/978-3-662-63202-4. ISBN: 9783662632017.

Klostermeier et al. 2019
Klostermeier, R.; Haag, S. & Benlian, A. (2019), *Geschäftsmodelle Digitaler Zwillinge,* Springer Fachmedien, Wiesbaden. ISBN: 9783658283520.

Kober et al. 2022
Kober, C.; Adomat, V.; Ahanpanjeh, M.; Fette, M. & Wulfsberg, J. P. (2022), „Digital Twin Fidelity Requirements Model for Manufacturing“. *3rd Conference on Production Systems and Logistics,* eds. D. Herberger & M. Hübner, pp. 595–611. https://doi.org/10.15488/12145.

Kotiades 2016
Kotiades, K. (2016), „Towards self-adaptive discrete event simulation (SADES)“ in *Operational Research Society Simulation Workshop 2016 (SW16),* eds. A. Anagnostou, K. Hoad & M. Kunc.

Kritzinger et al. 2018
Kritzinger, W.; Karner, M.; Traar, G.; Henjes, J. & Sihn, W. (2018), „Digital Twin in

manufacturing: A categorical literature review and classification“, *IFAC-PapersOnLine*, 51(11), pp. 1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474.

Krugh et al. 2017
Krugh, M.; McGee, E.; McGee, S.; Mears, L.; Ivanco, A.; Podd, K. C. & Watkins, B. (2017), „Measurement of Operator-machine Interaction on a Chaku-chaku Assembly Line“, *Procedia Manufacturing*, 10, pp. 123–135. https://doi.org/10.1016/j.promfg.2017.07.039.

Kuehner et al. 2021
Kuehner, K. J.; Scheer, R. & Strassburger, S. (2021), „Digital twin - finding common ground - a meta-review“, *Procedia CIRP*, 104, pp. 1227–1232. https://doi.org/10.1016/j.procir.2021.11.206.

Kuhn 2017
Kuhn, T. (2017), „Digitaler Zwilling“, *Informatik-Spektrum*, 40(5), pp. 440–444. https://doi.org/10.1007/s00287-017-1061-2.

Kuhnle et al. 2021
Kuhnle, A.; Kaiser, J.-P.; Theiß, F.; Stricker, N. & Lanza, G. (2021), „Designing an adaptive production control system using reinforcement learning“, *Journal of Intelligent Manufacturing*, 32(3), pp. 855–876. https://doi.org/10.1007/s10845-020-01612-y.

Kuhnle et al. 2019
Kuhnle, A.; Röhrig, N. & Lanza, G. (2019), „Autonomous order dispatching in the semiconductor industry using reinforcement learning“. *12th CIRP Conference on Intelligent Computation in Manufacturing Engineering,* eds. R. Teti & D. M. D’Addona, pp. 391–396. https://doi.org/10.1016/j.procir.2019.02.101.

Kuo & Kusiak 2019
Kuo, Y.-H. & Kusiak, A. (2019), „From data to big data in production research: the past and future trends“, *International Journal of Production Research*, 57(15-16), pp. 4828–4853. https://doi.org/10.1080/00207543.2018.1443230.

Kurbel 2013
Kurbel, K. E. (2013), *Enterprise resource planning and supply chain management. Functions, Business Processes and Software for Manufacturing Companies,* Springer, Berlin, Heidelberg. ISBN: 978-3-642-31572-5.

Landherr et al. 2013

Landherr, M.; Neumann, M.; Volkmann, J.; Jäger, J.; Kluth, A.; Lucke, D.; Rahman, O.-A.; Riexinger, G. & Constantinescu, C. (2013), „Fabriklebenszyklus-management“ in *Digitale Produktion,* Springer, Berlin, Heidelberg, pp. 163–195.

Law 2015

Law, A. M. (2015), *Simulation modeling and analysis,* McGraw-Hill Education, Dubuque. ISBN: 978-0-07-340132-4.

Law 2019

Law, A. M. (2019), „How to Build Valid and Credible Simulation Models“. *2019 Winter Simulation Conference (WSC),* eds. N. Mustafee, K.-H. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas & Y.-J. Son, IEEE, pp. 1402–1414. https://doi.org/10.1109/wsc40007.2019.9004789.

Law 2009

Law, A. M. (2009), „How to build valid and credible simulation models“. *Proceedings of the 2009 Winter Simulation Conference (WSC),* eds. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin & R. G. Ingalls, IEEE, pp. 24–33. ISBN: 978-1-4244-5770-0. https://doi.org/10.1109/WSC.2009.5429312.

Lechler et al. 2020

Lechler, T.; Fuchs, J.; Sjarov, M.; Brossog, M.; Selmaier, A.; Faltus, F.; Donhauser, T. & Franke, J. (2020), „Introduction of a comprehensive Structure Model for the Digital Twin in Manufacturing“. *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA),* eds. T. Sauter, Vasques, Francisco, L. Lo Bello & V. Vyatkin, IEEE, pp. 1773–1780. https://doi.org/10.1109/etfa46521.2020.9212030.

Lepenioti et al. 2020

Lepenioti, K.; Bousdekis, A.; Apostolou, D. & Mentzas, G. (2020), „Prescriptive analytics: Literature review and research challenges“, *International Journal of Information Management*, 50(3), pp. 57–70. https://doi.org/10.1016/j.ijinfomgt.2019.04.003.

Liu et al. 2021

Liu, M.; Fang, S.; Dong, H. & Xu, C. (2021), „Review of digital twin about concepts, technologies, and industrial applications“, *Journal of Manufacturing Systems*, 58, pp. 346–361. https://doi.org/10.1016/j.jmsy.2020.06.017.

Liu et al. 2018

Liu, Z.; Meyendorf, N. & Mrad, N. (2018), „The role of data fusion in predictive maintenance using digital twin“. *AIP Conference Proceedings. 44th Review of Progress in Quantitative Nondestructive Evaluation,* eds. L. J. Bond & D. E. Chimenti, 020023 - 1-6. ISBN: 978-0-7354-1644-4. https://doi.org/10.1063/1.5031520.

Magnanini & Tolio 2021

Magnanini, M. C. & Tolio, T. A.M. (2021), „A model-based Digital Twin to support responsive manufacturing systems“, *CIRP Annals*(70), pp. 1–4.

Manoury et al. 2021

Manoury, M.; Brünnhäußer, J.; Geiger, A.; Lindow, K. & Stark, R. (2021), *Modellierung- und Simulationsbedarfe der intelligenten Fabrik*. *Expertise des Forschungsbeirats der Plattform Industrie 4.0*. report, Munich

März 2020

März, L. (2020), „Einsatzmöglichkeiten von Simulation und Optimierung in der Planung der Automobilindustrie“ in *Ablaufsimulation in der Automobilindustrie,* Springer Vieweg, Berlin, Heidelberg, pp. 365–377.

Mayer et al. 2020

Mayer, G.; Pöge, C.; Spieckermann, S. & Wenzel, S. (eds.) (2020), *Ablaufsimulation in der Automobilindustrie,* Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN: 978-3-662-59387-5.

Meier & Kaufmann 2016

Meier, A. & Kaufmann, M. (2016), *SQL- & NoSQL-Datenbanken,* Springer, Berlin, Heidelberg. ISBN: 9783662476642.

Meudt et al. 2017

Meudt, T.; Pohl, M. & Metternich, J. (2017), *Die Automatisierungspyramide - Ein Literaturüberblick,* non-peer-reviewed university report, PTW CIP, TU Darmstadt, Darmstadt. https://tuprints.ulb.tu-darmstadt.de/6298/1/2017%20-%20Die%20Automatisierungspyramide%20-%20Ein%20Literatur%C3%BCberblick-2.pdf [February 1st, 2023]

Mieschner & Mayer 2020

Mieschner, M. & Mayer, G. (2020), „Bereitstellung und Verwaltung von Simulationseingangsdaten“ in *Ablaufsimulation in der Automobilindustrie,* Springer Vieweg, Berlin, Heidelberg, pp. 319–332.

Mieth et al. 2019

Mieth, C.; Meyer, A. & Henke, M. (2019), „Framework for the usage of data from real-time indoor localization systems to derive inputs for manufacturing simulation“, *Procedia CIRP*, 81(11), pp. 868–873. https://doi.org/10.1016/j.procir.2019.03.216.

Mourtzis 2020

Mourtzis, D. (2020), „Simulation in the design and operation of manufacturing systems: state of the art and new trends“, *International Journal of Production Research*, 58(7), pp. 1927–1949. https://doi.org/10.1080/00207543.2019.1636321.

Müller-Sommer 2010

Müller-Sommer, H. (20. Mai 2010), *Aufwände der Belieferungssimulation*. *Umfrage unter den Simulationsexperten der deutschen Automobilherstellern durch das Fachgebiet Wirtschaftsinformatik für Industriebetriebe der TU Ilmenau*. https://www.db-thueringen.de/servlets/MCRFileNodeServlet/dbt_derivate_00020567/Aufwand%20der%20Belieferungssimulation%20-%20Umfrage%20im%20VDA.pdf [15 April 2023]

Müller-Sommer 2013

Müller-Sommer, H. (2013), *Wirtschaftliche Generierung von Belieferungssimulationen unter Verwendung rechnerunterstützter Plausibilisierungsmethoden für die Bewertung der Eingangsdaten*. Dissertation, TU Ilmenau, Ilmenau.

Nagahara et al. 2020

Nagahara, S.; Serita, S.; Shiho, Y.; Zheng, S.; Wang, H.; Chida, T. & Gupta, C. (2020), „Toward data-driven modeling of material flow simulation: automatic parameter calibration of multiple agents from sparse production log“. *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE),* eds. unknown, IEEE. https://doi.org/10.1109/case48305.2020.9216832.

Nandimath et al. 2013

Nandimath, J.; Banerjee, E.; Patil, A.; Kakade, P.; Vaidya, S. & Chaturvedi, D. (2013), „Big data analysis using Apache Hadoop“. *2013 IEEE 14th International*

Conference on Information Reuse & Integration (IRI), eds. unknown, IEEE. https://doi.org/10.1109/iri.2013.6642536.

Negahban & Smith 2014

Negahban, A. & Smith, J. S. (2014), „Simulation for manufacturing system design and operation: Literature review and analysis", *Journal of Manufacturing Systems*, 33(2), pp. 241–261. https://doi.org/10.1016/j.jmsy.2013.12.007.

Negri et al. 2017

Negri, E.; Fumagalli, L. & Macchi, M. (2017), „A Review of the Roles of Digital Twin in CPS-based Production Systems", *Procedia Manufacturing*, 11, pp. 939–948. https://doi.org/10.1016/j.promfg.2017.07.198.

Neto et al. 2020

Neto, A. A.; Deschamps, F.; da Silva, E. R. & Lima, E. P. de (2020), „Digital twins in manufacturing: an assessment of drivers, enablers and barriers to implementation", *Procedia CIRP*, 93, pp. 210–215. https://doi.org/10.1016/j.procir.2020.04.131.

Onggo et al. 2020

Onggo, B. S.; Corlu, C. G.; Juan, A. A.; Monks, T. & La Torre, R. de (2020), „Combining symbiotic simulation systems with enterprise data storage systems for real-time decision-making", *Enterprise Information Systems*, 1(5), pp. 1–18. https://doi.org/10.1080/17517575.2020.1777587.

Onggo et al. 2010

Onggo, B. S.; Hill, J. & Brooks, R. (2010), „A pilot survey on data identification and collection in simulation projects". *Proceedings of the 27th European Simulation and Modelling Conference,* eds. W. Rekdalsbakken, R. T. B. Bye & H. Zhang. ISBN: 978-0-9564944-6-7.

Ottino 2003

Ottino, J. M. (2003), „Complex systems", *AIChE Journal*, 49(2), pp. 292–299. https://doi.org/10.1002/aic.690490202.

Overbeck et al. 2021a

Overbeck, L.; Brützel, O.; Teufel, M.; Stricker, S.; Kuhnle, A. & Lanza, G. (2021), „Continuous adaption through real data analysis turn simulation models into digital twins“, *Procedia CIRP*, 104, pp. 98–103. https://doi.org/10.1016/j.procir.2021.11.017.

Overbeck et al. 2020

Overbeck, L.; Brützel, O.; Stricker, S. & Lanza, G. (2020), „Digitaler Zwilling des Produktionssystems“, *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 115(Special), pp. 62–65. https://doi.org/10.3139/104.112326.

Overbeck et al. 2022

Overbeck, L.; Rose, A.; May, M. & Lanza, G. (2022), „Nutzungskonzept für Digitale Zwillinge von Produktionssystemen“, *Zeitschrift für wirtschaftlichen Fabrikbetrieb*, 117(4), pp. 244–248. https://doi.org/10.1515/zwf-2022-1035.

Overbeck et al. 2021b

Overbeck, L.; Hugues, A.; May, M. C.; Kuhnle, A. & Lanza, G. (2021), „Reinforcement Learning Based Production Control of Semi-automated Manufacturing Systems“, *Procedia CIRP*, 103(7676), pp. 170–175. https://doi.org/10.1016/j.procir.2021.10.027.

Overbeck et al. 2021

Overbeck, L.; Le Louarn, A.; Brützel, O.; Stricker, N. & Lanza, G. (2021), „Continuous Validation and Updating for High Accuracy of Digital Twins of Production Systems“. *Simulation in Produktion und Logistik 2021,* eds. J. Franke & P. Schuderer, Cuvillier Verlag, Göttingen, pp. 609–617.

Overbeck et al. 2023

Overbeck, L.; Graves, S. C. & Lanza, G. (2023), „ Development and analysis of digital twins of production systems “, *International Journal of Production Research*, https://doi.org/10.1080/00207543.2023.2242525

Paprotny et al. 1999

Paprotny, I.; Zhao, W. & Mackulak, G. T. (1999), „Reducing model creation cycle time by automated conversion of a CAD AHMS layout design“. *WSC’99. 1999 Winter Simulation Conference Proceedings. ‘Simulation - A Bridge to the Future’,* eds. P. A. Farrington, H. B. Nembhard, D. T. Sturrock & G. W. Evans, IEEE. https://doi.org/10.1109/wsc.1999.823259.

Petri 1962

Petri, C. A. (1962), *Kommunikation mit Automaten*. Dissertation, Universität Bonn, Bonn.

Priese & Wimmel 2008

Priese, L. & Wimmel, H. (2008), *Petri-Netze,* Springer-Verlag. Berlin, Heidelberg ISBN: 9783540769712.

Rabe et al. 2008

Rabe, M.; Spieckermann, S. & Wenzel, S. (2008), *Verifikation und Validierung für die Simulation in Produktion und Logistik*. *Vorgehensmodelle und Techniken,* Springer, Berlin, Heidelberg. ISBN: 9783540352815.

Razali & Yap 2011

Razali, N. M. & Yap, B. W. (2011), „Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests“, *Journal of Statistical Modeling and Analytics*, 2(1), pp. 21–33.

Redelinghuys et al. 2020

Redelinghuys, A. J. H.; Basson, A. H. & Kruger, K. (2020), „A six-layer architecture for the digital twin: a manufacturing case study implementation“, *Journal of Intelligent Manufacturing*, 31(6), pp. 1383–1402. https://doi.org/10.1007/s10845-019-01516-6.

Reinhardt et al. 2019

Reinhardt, H.; Weber, M. & Putz, M. (2019), „A Survey on Automatic Model Generation for Material Flow Simulation in Discrete Manufacturing“, *Procedia CIRP*, 81(2), pp. 121–126. https://doi.org/10.1016/j.procir.2019.03.022.

Repschläger et al. 2010

Repschläger, J.; Pannicke, D. & Zarnekow, R. (2010), „Cloud Computing: Definitionen, Geschäftsmodelle und Entwicklungspotenziale“, *HMD Praxis der Wirtschaftsinformatik*, 47(5), pp. 6–15. https://doi.org/10.1007/BF03340507.

Riedelsheimer et al. 2020

Riedelsheimer, T.; Lünnemann, P.; Wehking, S. & Dorfhuber, L. (2020), *Digital Twin Readiness Assessment*. *Eine Studie zum digitalen Zwilling in der fertigenden Industrie*. Report, Fraunhofer IPK, Berlin

Robertson & Perera 2002

Robertson, N. & Perera, T. (2002), „Automated data collection for simulation?“, *Simulation Practice and Theory*, 9, pp. 349–364. https://doi.org/10.1016/S0928-4869(01)00055-6.

Robinson et al. 2010

Robinson, S.; Brooks, R.; Kotiadis, K. & van der Zee, D.-J. (eds.) (2010), *Conceptual Modeling for Discrete-Event Simulation,* CRC Press. http://dx.doi.org/10.1201/9781439810385. ISBN: 9781439810385.

Robinson 1997

Robinson, S. (1997), „Simulation model verification and validation: Increasing the users’ confidence“. *1997 Winter Simulation Conference,* eds. S. Andradottir, K. J. Healy, D. H. Withers & B. L. Nelson, IEEE, pp. 53–59.

Robinson & Bhatia 1995

Robinson, S. & Bhatia, V. (1995), „Secrets of successful simulation projects“. *Proceedings of the 1995 Winter Simulation Conference,* eds. K. Alexopoulos, K. Kang, W. R. Lilegdon & D. Goldsman, IEEE, Piscataway, New Jersey, pp. 61–67. https://doi.org/10.1109/wsc.1995.478706.

Roddeck 2019

Roddeck, W. (2019), *Einführung in die Mechatronik,* Springer Vieweg, Wiesbaden. ISBN: 9783658277758.

Rooks 2009

Rooks, T. (2009), *Rechnergestützte Simulationsmodellgenerierung zur dynamischen Absicherung der Montagelogistikplanung bei der Fahrzeugneutypplanung im Rahmen der Digitalen Fabrik*. Dissertation, TU Clausthal, Clausthal.

Rose & März 2010

Rose, O. & März, L. (2010), „Simulation“ in *Simulation und Optimierung in Produktion und Logistik. Praxisorientierter Leitfaden mit Fallbeispielen,* eds. L. März, W. Krug, O. Rose & G. Weigert, Springer, Heidelberg Dordrecht London New York, pp. 13–19.

Rosen et al. 2015

Rosen, R.; Wichert, G. von; Lo, G. & Bettenhausen, K. D. (2015), „About The Importance of Autonomy and Digital Twins for the Future of Manufacturing“, *IFAC-PapersOnLine*, 48(3), pp. 567–572. https://doi.org/10.1016/j.ifacol.2015.06.141.

Rozinat et al. 2009

Rozinat, A.; Mans, R. S.; Song, M. & van der Aalst, W. M.P. (2009), „Discovering simulation models“, *Information Systems*, 34(3), pp. 305–327. https://doi.org/10.1016/j.is.2008.09.002.

Salam & Stevens 2006

Salam, A. F. & Stevens, J. (2006), *Semantic Web Technologies and E-Business: Toward the Integrated Virtual Organization and Business Process Automation*. *Toward the Integrated Virtual Organization and Business Process Automation,* Idea Group Inc (IGI). ISBN: 9781599041940.

Santos et al. 2022

Santos, C. H. d.; de Queiroz, J. A.; Leal, F. & Montevechi, J. A. B. (2022), „Use of simulation in the industry 4.0 context: Creation of a Digital Twin to optimise decision making on non-automated process“, *Journal of Simulation*, 16(3), pp. 284–297. https://doi.org/10.1080/17477778.2020.1811172.

Sargent 2020

Sargent, R. G. (2020), „Verification And Validation Of Simulation Models: An Advanced Tutorial“. *Proceedings of the Winter Simulation Conference,* eds. K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder & R. Thiesing, IEEE, pp. 16–29. ISBN: 978-1-7281-9499-8. https://doi.org/10.1109/WSC48552.2020.9384052.

Sargent 1991

Sargent, R. G. (1991), „Simulation model verification and validation“ in *Proceeding of the 1991 Winter Simulation Conference,* eds. B. L. Nelson, D. Kelton & G. M. Clark, pp. 37–47.

Sargent 2013

Sargent, R. G. (2013), „An introduction to verification and validation of simulation models“. *2013 Winter Simulations Conference (WSC),* eds. R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill & M. E. Kuhl, IEEE, pp. 321–327. ISBN: 978-1-4799-3950-3. https://doi.org/10.1109/WSC.2013.6721430.

Scheer et al. 2021

Scheer, R.; Straßburger, S. & Knapp, M. (2021), „Digital-physische Verbundkonzepte: Gegenüberstellung, Nutzeffekte und kritische Hürden“. *Simulation in Produktion und Logistik 2021,* eds. J. Franke & P. Schuderer, Cuvillier Verlag, Göttingen, pp. 11–20.

Schmidt-Volkmar 2008

Schmidt-Volkmar, P. (2008), *Betriebswirtschaftliche Analyse auf operationalen Daten,* Gabler, Wiesbaden. ISBN: 78-3-8349-1230-5.

Schwarz et al. 1978

Schwarz, L. B.; Graves, S. C. & Hausman, W. H. (1978), „Scheduling Policies for Automatic Warehousing Systems: Simulation Results“, *A I I E Transactions*, 10(3), pp. 260–270. https://doi.org/10.1080/05695557808975213.

Scitovski et al. 2021

Scitovski, R.; Sabo, K.; Martínez-Álvarez, F. & Ungar, Š. (2021), *Cluster Analysis and Applications,* Springer International Publishing, Cham. http://dx.doi.org/10.1007/978-3-030-74552-3. ISBN: 9783030745516.

Selke 2005

Selke, C. (2005), *Entwicklung von Methoden zur automatischen Simulationsmodellgenerierung*. Dissertation, TU München, München, iwb.

Shafto et al. 2010

Shafto, M.; Conroy, M.; Doyle, R.; Glaessgen, E.; Kemp, C.; LeMoigne, J. & Wang, L. (2010), *NASA Modeling, Simulation, Information Technology & Processing - TA11*.

Shao & Helu 2020

Shao, G. & Helu, M. (2020), „Framework for a digital twin in manufacturing: Scope and requirements“, *Manufacturing Letter*, 24, pp. 105–107.

Shehab et al. 2004

Shehab, E. M.; Sharp, M. W.; Supramaniam, L. & Spedding, T. A. (2004), „Enterprise resource planning: An integrative review“, *Business Process Management Journal*, 10(4), pp. 359–386. https://doi.org/10.1108/14637150410548056.

Siepmann 2016

Siepmann, D. (2016), „Industrie 4.0–Technologische Komponenten“ in *Einführung* *und Umsetzung von Industrie 4.0,* eds. A. Roth, Springer Gabler Verlag, Berlin, Heidelberg, pp. 47–72.

SISO 2010

SISO (2010), *Standard for: Core Manufacturing Simulation Data ─ UML Model*. SISO-STD-008-2010.

Sitz et al. 2021

Sitz, S.; Zerreis, M.; Lechler, T. & Franke, J. (2021), „Einsatz der ereignisdiskreten Materialflusssimulation bei Methoden aus der kontinuierlichen Verbesserung im Fertigungsumfeld: Erfolg und Auswirkungen der Speedweek 4.0". *Simulation in Produktion und Logistik 2021,* eds. J. Franke & P. Schuderer, Cuvillier Verlag, Göttingen, pp. 143–156. ISBN: 9783736974791.

Sjarov et al. 2020

Sjarov, M.; Lechler, T.; Fuchs, J.; Brossog, M.; Selmaier, A.; Faltus, F.; Donhauser, T. & Franke, J. (2020), „The Digital Twin Concept in Industry – A Review and Systematization". *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA),* eds. T. Sauter, Vasques, Francisco, L. Lo Bello & V. Vyatkin, IEEE, pp. 1789–1796. https://doi.org/10.1109/etfa46521.2020.9212089.

Skoogh et al. 2012

Skoogh, A.; Johansson, B. & Stahre, J. (2012), „Automated input data management: evaluation of a concept for reduced time consumption in discrete event simulation", *SIMULATION*, 88(11), pp. 1279–1293. https://doi.org/10.1177/0037549712443404.

Skoogh & Johansson 2007

Skoogh, A. & Johansson, B. (2007), „Time consumption analysis of input data activities in discrete event simulation projects". *Proceedings of the 2007 Swedish Production Symposium.,* eds. unknown.

Splanemann 1995

Splanemann, R. (1995), *Teilautomatische Generierung von Simulationsmodellen aus systemneutral definierten Unternehmensdaten*. Dissertation, Universität Bremen, Bremen, Institut für Betriebstechnik und angewandte Arbeitswissenschaften.

Stark et al. 2017

Stark, R.; Kind, S. & Neumeyer, S. (2017), „Innovations in digital modelling for next generation manufacturing system design“, *CIRP Annals*, 66(1), pp. 169–172. https://doi.org/10.1016/j.cirp.2017.04.045.

Stark & Damerau 2019

Stark, R. & Damerau, T. (2019), „Digital Twin“ in *CIRP Encyclopedia of Production Engineering,* eds. S. Chatti & T. Tolio, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–8.

Steiner 2021

Steiner, R. (2021), *Grundkurs relationale Datenbanken*. *Einführung in die Praxis der Datenbankentwicklung für Ausbildung, Studium und IT-Beruf,* Springer Vieweg, Wiesbaden. ISBN: 9783658328344.

Stjepandic et al. 2022

Stjepandic, J.; Sommer, M. & Denkena, B. (eds.) (2022), *DigiTwin: An Approach for Production Process Optimization in a Built Environment,* Springer, Cham. ISBN: 978-3-030-77538-4.

Tao et al. 2017

Tao, F.; Zhang, M.; Cheng, J. & Qi, Q. (2017), „Digital twin workshop: a new paradigm for future workshop“, *Computer Integrated Manufacturing Systems*, 23(1), pp. 1–9.

Tao & Zhang 2017

Tao, F. & Zhang, M. (2017), „Digital Twin Shop-Floor: A New Shop-Floor Paradigm Towards Smart Manufacturing“, *IEEE Access*, 5, pp. 20418–20427. https://doi.org/10.1109/access.2017.2756069.

Terkaj et al. 2021

Terkaj, W.; Qi, Q.; Urgo, M.; Scott, P. J. & Jiang, X. (2021), „Multi-scale modelling of manufacturing systems using ontologies and delta-lenses“, *CIRP Annals*, 70(1), pp. 361–364. https://doi.org/10.1016/j.cirp.2021.04.047.

Terkaj & Urgo 2015

Terkaj, W. & Urgo, M. (2015), „A Virtual Factory Data Model as a Support Tool for the Simulation of Manufacturing Systems“, *Procedia CIRP*, 28(1), pp. 137–142. https://doi.org/10.1016/j.procir.2015.04.023.

Thaker et al. 2004

Thaker, B. H.; Doebling, S. W.; Hernez, F. M.; Anderson, M. C.; Pepin, J. E. & Rodriguez, E. A. (2004), *Concepts of model verification and validation*. Los Alamos (USA)

Thiers et al. 2016

Thiers, G.; Sprock, T.; McGinnis, L.; Graunke, A. & Christian, M. (2016), „Automated production system simulations using commercial off-the-shelf simulation tools". *2016 Winter Simulation Conference (WSC),* eds. T.M.K. Roeder, P. I. Frazier, R. Szechtman, E. Zhou, T. Huschka & S. E. Chick, IEEE, Piscataway, NJ, pp. 1036–1047.

Uhlenkamp et al. 2019

Uhlenkamp, J.-F.; Hribernik, K.; Wellsandt, S. & Thoben, K.-D. (2019), „Digital Twin Applications : A first systemization of their dimensions" in *2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC) 2019*. http://dx.doi.org/10.1109/ice.2019.8792579.

Vachalek et al. 2017

Vachalek, J.; Bartalsky, L.; Rovny, O.; Sismisova, D.; Morhac, M. & Loksik, M. (2017), „The digital twin of an industrial production line within the industry 4.0 concept". *2017 21st International Conference on Process Control (PC),* IEEE, University of Miskolc, Hungary. ISBN: 978-963-358-132-2. https://doi.org/10.1109/pc.2017.7976223.

van der Aalst 2018

van der Aalst, W. M.P. (2018), „Simulation and Process Mining - Match made in Heaven!". *Proceedings of the 50th Computer Simulation Conference,* eds. unknown.

van der Aalst 2016

van der Aalst, W. M.P. (2016), *Process mining: data science in action,* Springer, Heidelberg. ISBN: 978-3-662-49850-7

van der Valk et al. 2022a

van der Valk, H.; Haße, H.; Möller, F. & Otto, B. (2022), „Archetypes of Digital Twins", *Business & Information Systems Engineering*, 64(3), pp. 375–391. https://doi.org/10.1007/s12599-021-00727-7.

van der Valk et al. 2022

van der Valk, H.; Henning, J.-L.; Winkelmann, S. & Haße, H. (2022), „Practical Requirements For Digital Twins In Production And Logistics“. *Conference of production systems and logistics 2022,* eds. D. Herberger & M. Hübner, publish-Ing., 32-41. https://doi.org/10.15488/12178.

van der Valk et al. 2020

van der Valk, H.; Haße, H.; Möller, F.; Arbter, M. & Otto, B. (2020), „A Taxonomy of Digital Twins“. *26th Americas Conference on Information Systems (AMCIS),* eds. B. Anderson, J. Thatcher & R. Meservy.

van der Valk et al. 2020

van der Valk, H.; Hunker, J.; Rabe, M. & Otto, B. (2020), „Digital Twins in Simulative Applications: A Taxonomy“. *2020 Winter Simulation Conference (WSC),* eds. K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. M.K. Roeder & R. Thiesing, IEEE, Piscataway, NJ. https://doi.org/10.1109/wsc48552.2020.9384051.

Vavilapalli et al. 2013

Vavilapalli, V. K.; Murthy, A. C.; Douglas, C.; Agarwal, S.; Konar, M.; Evans, R.; Graves, T.; Lowe, J.; Shah, H.; Seth, S.; Saha, B.; Curino, C.; O'Malley, O.; Radia, S.; Reed, B. & Baldeschwieler, E. (2013), „Apache hadoop yarn: Yet another resource negotiator“. *Proceedings of the 4th annual Symposium on Cloud Computing,* eds. unknown. https://doi.org/10.1145/2523616.2523633.

VDI 2014

VDI (2014), *VDI-Richtlinie 3633,* Beuth, Berlin, 03.100.10.

VDI 2016

VDI (2016), *VDI-Richtlinie 5600,* Beuth, Berlin, 35.240.50.

Vieira et al. 2018

Vieira, A. A.; Dias, L. M. S.; Santos, M. Y.; Pereira, G. A. B. & Oliveira, J. A. (2018), „Setting an industry 4.0 research and development agenda for simulation - a literature review“, *International Journal of Simulation Modelling*, 17(3), pp. 377–390. https://doi.org/10.2507/IJSIMM17(3)429.

Wellenreuther & Zastrow 2015

Wellenreuther, G. & Zastrow, D. (2015), *Automatisieren mit SPS - Theorie und Praxis,* Springer Vieweg, Wiesbaden. ISBN: 3834825972.

Wenzel 2008

Wenzel, S. (2008), „Modellbildung und Simulation in der Logistik“ in *Modellbildung und Simulation, Operations Research,* eds. A. Lehmann & T. Krieger, München.

Wenzel et al. 2008

Wenzel, S.; Weiß, M.; Collisi-Böhmer, S.; Pitsch, H. & Rose, O. (2008), *Qualitätskriterien für die Simulation in Produktion und Logistik*. *Planung und Durchführung von Simulationsstudie,* Springer, Berlin, Heidelberg. ISBN: 978-3-540-35272-3.

Wenzel 2018

Wenzel, S. (2018), „Simulation logistischer Systeme“ in *Modellierung logistischer Systeme,* eds. H. Tempelmeier, Springer Vieweg, Berlin, Heidelberg, pp. 1–34.

Werner & Weigert 2002

Werner, S. & Weigert, G. (2002), „Process accompanying simulation - a general approach for the continuous optimization of manufacturing schedules in electronics production“. *Proceedings of the Winter Simulation Conference,* eds. E. Yücesan, R. Chbeir, J. L. Snowdon & J. M. Charnes, IEEE, pp. 1903–1908. ISBN: 0-7803-7614-5. https://doi.org/10.1109/WSC.2002.1166487.

Westkämper 2007

Westkämper, E. (2007), „Digital Manufacturing In The Global Era“ in *Digital Enterprise Technology,* eds. P. F. Cunha & P. G. Maropoulos, Springer, Boston, MA, pp. 3–14.

Wiendahl et al. 2007

Wiendahl, H.-P.; ElMaraghy, H. A.; Nyhuis, P.; Zäh, M. F.; Wiendahl, H.-H.; Duffie, N. & Brieke, M. (2007), „Changeable Manufacturing - Classification, Design and Operation“, *CIRP Annals*, 56(2), pp. 783–809.
https://doi.org/10.1016/j.cirp.2007.10.003.

Wy et al. 2011

Wy, J.; Jeong, S.; Kim, B.-I.; Park, J.; Shin, J.; Yoon, H. & Lee, S. (2011), „A data-driven generic simulation model for logistics-embedded assembly manufacturing lines“, *Computers & Industrial Engineering*, 60(1), pp. 138–147.
https://doi.org/10.1016/j.cie.2010.10.011.

Zenner 2006

Zenner, C. (2006), *Durchgängiges Variantenmanagement in der Technischen Produktionsplanung*. Dissertation, Universität des Saarlandes, Saarbrücken. https://doi.org/10.22028/D291-22342.

Zhang & Deuse 2009

Zhang, W. & Deuse, J. (2009), „Cell staffing and standardized work design in Chaku-Chaku production lines using a hybrid optimization algorithm“. *2009 International Conference on Computers & Industrial Engineering,* eds. Dessouly M.I. & I. Kacem, IEEE, pp. 305–310. https://doi.org/10.1109/iccie.2009.5223901.

Zülch 2019

Zülch, G. (2019), „Entwicklung und Stand der Personaleinsatzsimulation und deren Einbeziehung in die Digitale Fabrik“ in *Simulation in Produktion und Logistik 2019,* eds. M. Putz & A. Schlegel, Wissenschaftliche Scripten, Auerbach.

List of Figures

Figure 1-1 Use of conventional simulation models versus the use of a Digital Twin 2

Figure 2-1 Digital Twins as the next evolutionary step in simulation technology entails its integration into the running systems (Rosen et al. 2015) 6

Figure 2-2 Components of the Digital Twin according to Stark et al. (2017) 7

Figure 2-3 Ways to examine a system (based on Law (2015, p. 4)) 11

Figure 2-4 Classification of simulation models according to time and state set. (Gutenschwager et al. 2017, p. 16) (based on Ören 1979, p. 36) 14

Figure 2-5 Sequence of a DES program (Law 2015, p.10) 16

Figure 2-6 Procedure model for simulation projects according to VDI 3633 (2014) 17

Figure 2-7 Four ways to enter data into the simulation model (Robertson & Perera 2002) 20

Figure 2-8 Possible Digital Twins on the structuring levels of production, based on (Westkämper 2007, p. 11; Wiendahl et al. 2007) 21

Figure 2-9 Hierarchical structure of a process based on Becker (2018, p. 9) 22

Figure 2-10 Exemplary Petri net and its elements 23

Figure 2-11 Classification of simulation tools for production and logistics, actualized and based on Eley (2012, p. 10) and Wenzel (2008) 27

Figure 2-12 The automation pyramid (Siepmann 2016) 30

Figure 2-13 Visualization clustering with the DBSCAN algorithm 37

Figure 2-14 The alpha algorithm 39

Figure 3-1 Structure model for Digital Twins according to Lechler et al. (2020) 41

Figure 4-1 Function sketch of the Digital Twin of the production system including the respective text sections 60

Figure 4-2 Choices to make during seed model creation 62

Figure 4-3 Validation process and previous settings 65

Figure 4-4 Significance of the explicit consideration of exceptional events in validation 68

Figure 4-5 Overview of the different options for validation of the Digital Twin 71

Figure 4-6 Examples of output curves 73

Figure 4-7 Illustration of the NRMSE calculation in the industrial use case 73

Figure 4-8 Examples of regression analysis with associated metrics 74

Figure 4-9 Exemplary output of the box plots 75

Figure 4-10 Possibility of considering several validation periods 77

Figure 4-11 Update process and prior settings 78

Figure 4-12 General procedure for determining the simulation input information 80

Figure 4-13 Structure of simulation input calculation 81

Figure 4-14 Histogram with four different distribution functions. The lognormal distribution fits best (A_Haizmann 2020) 83

Figure 4-15 Sequence of material flow detection by means of process mining 84

Figure 4-16 Schematic flow of the determination of the work processes in the production system 88

Figure 4-17 Considerations of changes in the production system 89

Figure 4-18 Schematic detection of commissioning and end of operation of machines A-D in line X 90

Figure 5-1 Schematic representation of the assembly and testing & completion subsystems 94

Figure 5-2 Model structure of Bosch use case based on (A_Nagel 2020) 97

Figure 5-3 Model view in Plant Simulation (2D) 97

Figure 5-4 Central data management in the simulation with update methods 98

Figure 5-5 Schematic of the standard station 99

Figure 5-6 Example of employee loops in generic production area with machines A - I 100

Figure 5-7 Single line in the software Plant Simulation 102

Figure 5-8 Overview of all four production lines in Plant Simulation 102

Figure 5-9 Sequence of validation based on A_Le Louarn (2021) 103

Figure 5-10 Distribution of failure durations of the automatically recorded failure messages 108

Figure 5-11 Distribution of durations of manually documented exceptional events 108

Figure 5-12 Update sequence used in Bosch use case 113

Figure 5-13 Web application with a scatter plot for a product variant at a station and fitted normal distribution according to A_Haizmann (2020, p. 77) 115

Figure 5-14 Example of distribution fitting at machines 11 and 13 for one product variant 116

Figure 5-15 Identification of pseudo failures by comparison with machine protocol 117

Figure 5-16 Material flow visualization with (a) and without (b) filter 121

Figure 5-17 Section of the Petri net of the material flow of a product variant as a result of process mining 121

Figure 5-18 Examples of observed changes in lines 1-3 at the machine level 123

Figure 5-19 Examples of observed changes in the production process 125

Figure 5-20 One possible configuration of the learning factory global production (Source: wbk) 126

Figure 5-21 Visualization of recorded worker movements in the learning factory 129

Figure 5-22 Employees' work areas identified with the aid of cluster analysis (A_Kudlik 2022) 130

Figure 5-23 Recognized layout and employee assignment in the learning factory in the U-layout (A_Kudlik 2022) 131

Figure 6-1 Exemplary periods for validation 133

Figure 6-2 Development of the accuracy of the Digital Twin of line 1 with and without update 134

Figure 6-3 Exemplary cumulated output curves for three weeks 135

Figure 6-4 Boxplots of OEE per hour in reality and simulation 135

Figure 6-5 Example scatter plot and regression metrics 135

Figure 6-6 Accuracy with different update period lengths 136

Figure 6-7 Accuracy depending on time since last update 137

Figure 6-8 Schema of sensitivity analysis 139

Figure 6-9 Change in relative error with variation of mean and standard deviation including the 95% confidence interval 140

Figure 6-10 Change of NRMSE when varying mean and standard deviation including the 95% confidence interval 141

Figure 6-11 Change of average NRMSE when varying the mean value including the 95% confidence interval 142

Figure 6-12 Change of average NRMSE when varying the standard deviation including the 95% confidence interval 143

Figure 6-13 Change of relative error when changing distribution type 144

Figure 6-14 NRMSE when machine availabilities are varied 145

Figure 6-15 NRMSE when MTTR are varied 146

Figure 6-16 Effect of changes in material flow data on accuracy of the Digital Twin 147

Figure 6-17 Influence of the information components on the accuracy of the Digital Twin for one exemplary week 148

Figure 6-18 Sequence for the creation of the user concept (Overbeck et al. 2022) 149

Figure 6-19 User concept Bosch use case (Overbeck et al. 2022) 150

Figure 6-20 Usage of Digital Twin for analyses and improvement projects over time 151

Figure 6-21 Determination of threshold to inform worker when to change subsystem 152

Figure 6-22 Example for improved worker loops in generic example 153

Figure 6-23 Influence of failures on different machines on the performance of the production system 154

Figure 6-24 Influence of the most important failure types of machine 7 on the performance of the production system 154

Figure 6-25 Example for the evaluation of hardware changes in the production system 155

Figure 6-26 Visualization of safety distance between worker during corona pandemic 156

List of Tables

Table 2-1 Classification of validation techniques with examples (Sargent 2020) 26

Table 2-2 Comparison of data warehouse and data lake (Khine & Wang 2018, p. 6) 33

Table 2-3 Example Event Log 38

Table 3-1 State of research 57

Table 4-1 Overview of metrics for the validation linked to regression analysis based on Bamberg (2017, p. 40) 74

Table 5-1 Example of OpenRequest list (A_Janikovits 2020, p. 68) 100

Table 5-2 Example of experimental investigation of possible limits for determining availability and exceptional events (EE) 109

Table 5-3 Periods considered for determining the updated input 112

Table 5-4 Example results of scrap calculation using SQL from data lake 118

Table 5-5 Example result of the filter process as preparation for material flow derivation by means of process mining. 119

Table 5-6 Example of the distribution of number of process steps in the use case 119

Table 5-7 Example of the frequency distribution of start and end processes 120

Table 5-8 Example routing table for the material flow of two product variants for the simulation model from the Petri net 122

Table 5-9 Assignment of employees to machines 130

Table 6-1 Initial accuracy of the Digital Twin of line 1 133

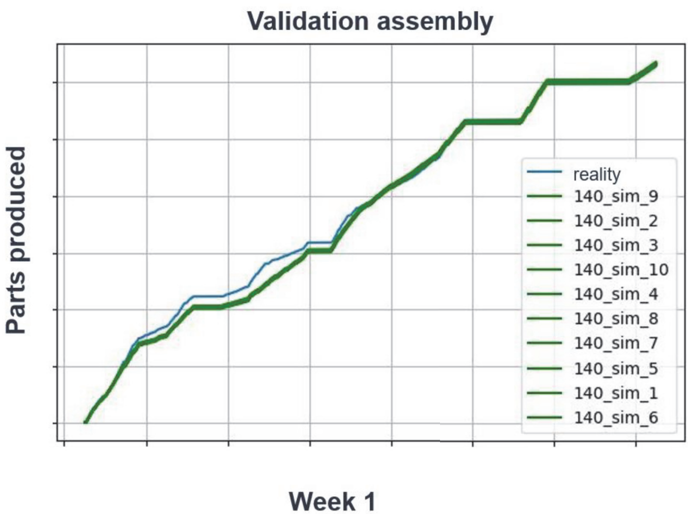
Table 6-2 Accuracy of the Digital Twin if update and validation period are identical 138

Table 7-1 Fulfillment of defined requirements by the presented approach 157

Appendix

A1 Example of detailed results of sensitivity analysis

Example for results of validation (one configuration of sensitivity analysis)				
Sim_ID	**Relative error assembly**	**NRMSD assembly**	**Relative error testing**	**NRMSD testing**
sim_1	0,02	4,29	0,11	3,84
sim_2	0,29	4,04	0,21	3,67
sim_3	0,29	4,27	0,24	4,29
sim_4	0,45	3,91	0,25	4,07
sim_5	0,54	4,51	0,30	3,93
sim_6	0,73	4,24	0,52	4,62
sim_7	0,78	3,84	0,54	3,96
sim_8	0,86	3,63	0,68	4,22
sim_9	0,92	3,91	0,85	4,54
sim_10	1,19	3,88	1,07	4,64
Average	**0,61**	**4,05**	**0,48**	**4,18**
Standard deviation	**0,34**	**0,25**	**0,30**	**0,32**
95% confidence interval	**0,21**	**0,16**	**0,19**	**0,20**



A2 User interface for validation and update of the Digital Twin

The user interface shown in Figure A2 is based on the existing visualization of the Plant Simulation software. It is supplemented by dialog boxes that allow the user to initiate specific updates or validations, make parameter selections, and set periods for validation and updates. In the case that the validation result is still negative after the execution of all possible automatic updates, the user interface informs the user about further possible adjustments that can only be carried out manually (for example changes in the number of machines).

In addition to running the validation and update as a fully automated closed-loop (the actual Digital Twin mechanism), each step can be triggered manually in the interface on its own, to allow for more in-depth analysis and testing of the Digital Twin. The coloring should only underline the different sections of the interface and has no deeper meaning.

Thus, a desired validation period can be set manually in the InputData dialog (1), the necessary information about the validation period can be queried and imported, and the corresponding simulation runs can be started. On the interface, the numerous pieces of information required for the model update are stored in tables, e.g. ‘Article_occupancy_changes’ (2) and “unplannedBreaks’ (3).

After running the simulation runs, the "Run validation" button (4) can be used to manually start the accuracy metric calculation, first exporting the simulation results and then triggering the Python script that queries the real comparison data and calculates the accuracy metrics.

After the validation feedback (in particular, whether it is successful or not), the update can be started via the "Run Update" button (5). This information is stored in the ‘UpdateInput’ table (6). By selecting the ‘ClosedLooped’ option (7), the automatic mutual triggering of validation and update can be activated, which constitutes the real automated Digital Twin process.

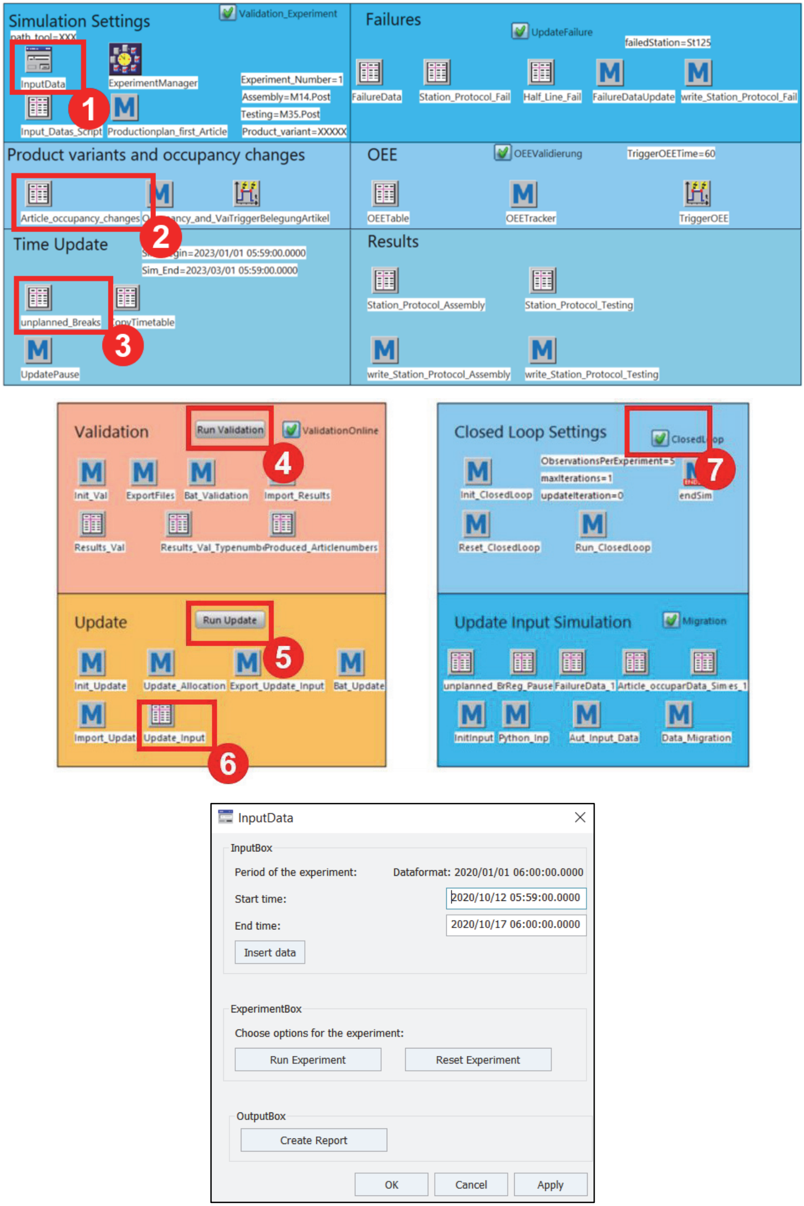


Figure A-2 User interface to manage validation and updating

A3 Examples of input tables for simulation model in Digital Twin

Schemata of data table of process times of automated process steps in simulation model for one line:

Product variant	Machine	Distribution type	My	Sigma	Lower Bound	Upper Bound
A	1	Lognormal	XXX	XXX	0.01	X
A	2	Normal	XXX	XXX	0.01	XXX
…	…	…	…	…	…	…
B	1	Lognormal	XXX	XXX	0.01	XXX

Schemata of data table of OK rates in simulation model for one line:

Product variant	Machine	OK rate
A	1	XXX
A	2	XXX
…	…	…
B	1	XXX

Schemata of data table machine availabilities in simulation model for one line:

Machine	Process	Failure	Availability	MTTR	Failure code	Failure description
1	Machine	Failure1	XXX	XXX	XXX	XXX
1	Machine	Failure2	XXX	XXX	XXX	XXX
…	…	…	…	…	…	…
2	Machine	Failure1	XXX	XXX	XXX	XXX

Schemata of data table of worker loops in simulation model for one line:

Number of workers	Product variant	Worker	Loop-index	location
n	A	1	1	Machine1.Pre
n	A	1	2	Machine1.Machine
n	A	1	3	Machine1.Post
…	…	…	…	…
n-1	A	1		

Forschungsberichte aus dem wbk
Institut für Produktionstechnik
Karlsruher Institut für Technologie (KIT)

Bisher erschienene Bände:

Band 0
Dr.-Ing. Wu Hong-qi

Adaptive Volumenstromregelung mit Hilfe von drehzahlgeregelten Elektroantrieben

Band 1
Dr.-Ing. Heinrich Weiß

Fräsen mit Schneidkeramik - Verhalten des System Werkzeugmaschine-Werkzeug-Werkstück und Prozessanalyse

Band 2
Dr.-Ing. Hans-Jürgen Stierle

Entwicklung und Untersuchung hydrostatischer Lager für die Axialkolbenmaschine

Band 3
Dr.-Ing. Herbert Hörner

Untersuchung des Geräuschverhaltens druckgeregelter Axialkolbenpumpen

Band 4
Dr.-Ing. Rolf-Dieter Brückbauer

Digitale Drehzahlregelung unter der besonderen Berücksichtigung von Quantisierungseffekten

Band 5
Dr.-Ing. Gerhard Staiger

Graphisch interaktive NC-Programmierung von Drehteilen im Werkstattbereich

Band 6
Dr.-Ing. Karl Peters

Ein Beitrag zur Berechnung und Kompensation von Positionierfehlern an Industrierobotern

Band 7
Dr.-Ing. Paul Stauss

Automatisierte Inbetriebnahme und Sicherung der Zuverlässigkeit und Verfügbarkeit numerisch gesteuerter Fertigungseinrichtungen

Band 8
Dr.-Ing. Günter Möckesch

Konzeption und Realisierung eines strategischen, integrierten Gesamtplanungs- und -bearbeitungssystems zur Optimierung der Drehteilorganisation für auftragsbezogene Drehereien

Band 9
Dr.-Ing. Thomas Oestreicher

Rechnergestützte Projektierung von Steuerungen

Band 10
Dr.-Ing. Thomas Selinger

Teilautomatisierte werkstattnahe NC-Programmerstellung im Umfeld einer integrierten Informationsverarbeitung

Band 11
Dr.-Ing. Thomas Buchholz

Prozessmodell Fräsen, Rechnerunterstützte Analyse, Optimierung und Überwachung

Band 12
Dr.-Ing. Bernhard Reichling

Lasergestützte Positions- und Bahnvermessung von Industrierobotern

Band 13
Dr.-Ing. Hans-Jürgen Lesser

Rechnergestützte Methoden zur Auswahl anforderungsgerechter Verbindungselemente

Band 14
Dr.-Ing. Hans-Jürgen Lauffer

Einsatz von Prozessmodellen zur rechnerunterstützten Auslegung von Räumwerkzeugen

Band 15
Dr.-Ing. Michael C. Wilhelm

Rechnergestützte Prüfplanung im Informationsverbund moderner Produktionssysteme

Band 16
Dr.-Ing. Martin Ochs

Entwurf eines Programmsystems zur wissensbasierten Planung und Konfigurierung

Band 17
Dr.-Ing. Heinz-Joachim Schneider

Erhöhung der Verfügbarkeit von hochautomatisierten Produktionseinrichtungen mit Hilfe der Fertigungsleittechnik

Band 18
Dr.-Ing. Hans-Reiner Ludwig

Beanspruchungsanalyse der Werkzeugschneiden beim Stirnplanfräsen

Band 19
Dr.-Ing. Rudolf Wieser

Methoden zur rechnergestützten Konfigurierung von Fertigungsanlagen

Band 20
Dr.-Ing. Edgar Schmitt

Werkstattsteuerung bei wechselnder Auftragsstruktur

Band 21
Dr.-Ing. Wilhelm Enderle

Verfügbarkeitssteigerung automatisierter Montagesysteme durch selbsttätige Behebung prozessbedingter Störungen

Band 22
Dr.-Ing. Dieter Buchberger

Rechnergestützte Strukturplanung von Produktionssystemen

Band 23
Prof. Dr.-Ing. Jürgen Fleischer

Rechnerunterstützte Technologieplanung für die flexibel automatisierte Fertigung von Abkantteilen

Band 24
Dr.-Ing. Lukas Loeffler

Adaptierbare und adaptive Benutzerschnittstellen

Band 25
Dr.-Ing. Thomas Friedmann

Integration von Produktentwicklung und Montageplanung durch neue rechnergestützte Verfahren

Band 26
Dr.-Ing. Robert Zurrin

Variables Formhonen durch rechnergestützte Hornprozesssteuerung

Band 27
Dr.-Ing. Karl-Heinz Bergen

Langhub-Innenrundhonen von Grauguss und Stahl mit einem elektromechanischem Vorschubsystem

Band 28
Dr.-Ing. Andreas Liebisch

Einflüsse des Festwalzens auf die Eigenspannungsverteilung und die Dauerfestigkeit einsatzgehärteter Zahnräder

Band 29
Dr.-Ing. Rolf Ziegler

Auslegung und Optimierung schneller Servopumpen

Band 30
Dr.-Ing. Rainer Bartl

Datenmodellgestützte Wissensverarbeitung zur Diagnose und Informationsunterstützung in technischen Systemen

Band 31
Dr.-Ing. Ulrich Golz

Analyse, Modellbildung und Optimierung des Betriebsverhaltens von Kugelgewindetrieben

Band 32
Dr.-Ing. Stephan Timmermann

Automatisierung der Feinbearbeitung in der Fertigung von Hohlformwerkzeugen

Band 33
Dr.-Ing. Thomas Noe

Rechnergestützter Wissenserwerb zur Erstellung von Überwachungs- und Diagnoseexpertensystemen für hydraulische Anlagen

Band 34
Dr.-Ing. Ralf Lenschow

Rechnerintegrierte Erstellung und Verifikation von Steuerungsprogrammen als Komponente einer durchgängigen Planungsmethodik

Band 35
Dr.-Ing. Matthias Kallabis

Räumen gehärteter Werkstoffe mit kristallinen Hartstoffen

Band 36
Dr.-Ing. Heiner-Michael Honeck

Rückführung von Fertigungsdaten zur Unterstützung einer fertigungsgerechten Konstruktion

Band 37
Dr.-Ing. Manfred Rohr

Automatisierte Technologieplanung am Beispiel der Komplettbearbeitung auf Dreh-/Fräszellen

Band 38
Dr.-Ing. Martin Steuer

Entwicklung von Softwarewerkzeugen zur wissensbasierten Inbetriebnahme von komplexen Serienmaschinen

Band 39
Dr.-Ing. Siegfried Beichter

Rechnergestützte technische Problemlösung bei der Angebotserstellung von flexiblen Drehzellen

Band 40
Dr.-Ing. Thomas Steitz

Methodik zur marktorientierten Entwicklung von Werkzeugmaschinen mit Integration von funktionsbasierter Strukturierung und Kostenschätzung

Band 41
Dr.-Ing. Michael Richter

Wissensbasierte Projektierung elektrohydraulischer Regelungen

Band 42
Dr.-Ing. Roman Kuhn

Technologieplanungssystem Fräsen. Wissensbasierte Auswahl von Werkzeugen, Schneidkörpern und Schnittbedingungen für das Fertigingsverfahren Fräsen

Band 43
Dr.-Ing. Hubert Klein

Rechnerunterstützte Qualitätssicherung bei der Produktion von Bauteilen mit frei geformten Oberflächen

Band 44
Dr.-Ing. Christian Hoffmann

Konzeption und Realisierung eines fertigungsintegrierten Koordinatenmessgerätes

Band 45
Dr.-Ing. Volker Frey

Planung der Leittechnik für flexible Fertigungsanlagen

Band 46
Dr.-Ing. Achim Feller

Kalkulation in der Angebotsphase mit dem selbsttätig abgeleiteten Erfahrungswissen der Arbeitsplanung

Band 47
Dr.-Ing. Markus Klaiber

Produktivitätssteigerung durch rechnerunterstütztes Einfahren von NC-Programmen

Band 48
Dr.-Ing. Roland Minges

Verbesserung der Genauigkeit beim fünfachsigen Fräsen von Freiformflächen

Band 49
Dr.-Ing. Wolfgang Bernhart

Beitrag zur Bewertung von Montagevarianten: Rechnergestützte Hilfsmittel zur kostenorientierten, parallelen Entwicklung von Produkt und Montagesystem

Band 50
Dr.-Ing. Peter Ganghoff

Wissensbasierte Unterstützung der Planung technischer Systeme: Konzeption eines Planungswerkzeuges und exemplarische Anwendung im Bereich der Montagesystemplanung

Band 51
Dr.-Ing. Frank Maier

Rechnergestützte Prozessregelung beim flexiblen Gesenkbiegen durch Rückführung von Qualitätsinformationen

Band 52
Dr.-Ing. Frank Debus

Ansatz eines rechnerunterstützten Planungsmanagements für die Planung in verteilten Strukturen

Band 53
Dr.-Ing. Joachim Weinbrecht

Ein Verfahren zur zielorientierten Reaktion auf Planabweichungen in der Werkstattregelung

Band 54
Dr.-Ing. Gerd Herrmann

Reduzierung des Entwicklungsaufwandes für anwendungsspezifische Zellenrechnersoftware durch Rechnerunterstützung

Band 55
Dr.-Ing. Robert Wassmer

Verschleissentwicklung im tribologischen System Fräsen: Beiträge zur Methodik der Prozessmodellierung auf der Basis tribologisher Untersuchungen beim Fräsen

Band 56
Dr.-Ing. Peter Uebelhoer

Inprocess-Geometriemessung beim Honen

Band 57
Dr.-Ing. Hans-Joachim Schelberg

Objektorientierte Projektierung von SPS-Software

Band 58
Dr.-Ing. Klaus Boes

Integration der Qualitätsentwicklung in featurebasierte CAD/CAM-Prozessketten

Band 59
Dr.-Ing. Martin Schreiber

Wirtschaftliche Investitionsbewertung komplexer Produktionssysteme unter Berücksichtigung von Unsicherheit

Band 60
Dr.-Ing. Ralf Steuernagel

Offenes adaptives Engineering-Werkzeug zur automatisierten Erstellung von entscheidungsunterstützenden Informationssystemen

Band 62
Dr.-Ing. Uwe Schauer

Qualitätsorientierte Feinbearbeitung mit Industrierobotern: Regelungsansatz für die Freiformflächenfertigung des Werkzeug- und Formenbaus

Band 63
Dr.-Ing. Simone Loeper

Kennzahlengestütztes Beratungssystem zur Verbesserung der Logistikleistung in der Werkstattfertigung

Band 64
Dr.-Ing. Achim Raab

Räumen mit hartstoffbeschichteten HSS-Werkzeugen

Band 65,
Dr.-Ing. Jan Erik Burghardt

Unterstützung der NC-Verfahrenskette durch ein bearbeitungs-elementorientiertes, lernfähiges Technologieplanungssystem

Band 66
Dr.-Ing. Christian Tritsch

Flexible Demontage technischer Gebrauchsgüter: Ansatz zur Planung und (teil-)automatisierten Durchführung industireller Demontageprozesse

Band 67
Dr.-Ing. Oliver Eitrich

Prozessorientiertes Kostenmodell für die entwicklungsbegleitende Vorkalkulation

Band 68
Dr.-Ing. Oliver Wilke

Optimierte Antriebskonzepte für Räummaschinen - Potentiale zur Leistungs-steigerung

Band 69
Dr.-Ing. Thilo Sieth

Rechnergestützte Modellierungsmethodik zerspantechnologischer Prozesse

Band 70
Dr.-Ing. Jan Linnenbuerger

Entwicklung neuer Verfahren zur automatisierten Erfassung der geometri-schen Abweichungen an Linearachsen und Drehschwenkköpfen

Band 71
Dr.-Ing. Mathias Klimmek

Fraktionierung technischer Produkte mittels eines frei beweglichen Wasserstrahlwerkzeuges

Band 72
Dr.-Ing. Marko Hartel

Kennzahlenbasiertes Bewertungssystem zur Beurteilung der Demontage- und Recyclingeignung von Produkten

Band 73
Dr.-Ing. Jörg Schaupp

Wechselwirkung zwischen der Maschinen- und Hauptspindelantriebsdynamik und dem Zerspanprozess beim Fräsen

Band 74
Dr.-Ing. Bernhard Neisius

Konzeption und Realisierung eines experimentellen Telemanipulators für die Laparoskopie

Band 75
Dr.-Ing. Wolfgang Walter

Erfolgsversprechende Muster für betriebliche Ideenfindungsprozesse. Ein Beitrag zur Steigerung der Innovationsfähigkeit

Band 76
Dr.-Ing. Julian Weber

Ein Ansatz zur Bewertung von Entwicklungsergebnissen in virtuellen Szenarien

Band 77
Dr.-Ing. Dipl. Wirtsch.-Ing. Markus Posur

Unterstützung der Auftragsdurchsetzung in der Fertigung durch Kommunikation über mobile Rechner

Band 78
Dr.-Ing. Frank Fleissner

Prozessorientierte Prüfplanung auf Basis von Bearbeitungsobjekten für die Kleinserienfertigung am Beispiel der Bohr- und Fräsbearbeitung

Band 79
Dr.-Ing. Anton Haberkern

Leistungsfähigere Kugelgewindetriebe durch Beschichtung

Band 80
Dr.-Ing. Dominik Matt

Objektorientierte Prozess- und Strukturinnovation (OPUS)

Band 81
Dr.-Ing. Jürgen Andres

Robotersysteme für den Wohnungsbau: Beitrag zur Automatisierung des Mauerwerkabaus und der Elektroinstallation auf Baustellen

Band 82
Dr.-Ing. Dipl.Wirtschaftsing. Simone Riedmiller

Der Prozesskalender - Eine Methodik zur marktorientierten Entwicklung von Prozessen

Band 83
Dr.-Ing. Dietmar Tilch

Analyse der Geometrieparameter von Präzisionsgewinden auf der Basis einer Least-Squares-Estimation

Band 84
Dr.-Ing. Dipl.-Kfm. Oliver Stiefbold

Konzeption eines reaktionsschnellen Planungssystems für Logistikketten auf Basis von Software-Agenten

Band 85
Dr.-Ing. Ulrich Walter

Einfluss von Kühlschmierstoff auf den Zerspanprozess beim Fräsen: Beitrag zum Prozessverständniss auf Basis von zerspantechnischen Untersuchungen

Band 86
Dr.-Ing. Bernd Werner

Konzeption von teilautonomer Gruppenarbeit unter Berücksichtigung kultureller Einflüsse

Band 87
Dr.-Ing. Ulf Osmers

Projektieren Speicherprogrammierbarer Steuerungen mit Virtual Reality

Band 88
Dr.-Ing. Oliver Doerfel

Optimierung der Zerspantechnik beim Fertigungsverfahren Wälzstossen: Analyse des Potentials zur Trockenbearbeitung

Band 89
Dr.-Ing. Peter Baumgartner

Stufenmethode zur Schnittstellengestaltung in der internationalen Produktion

Band 90
Dr.-Ing. Dirk Vossmann

Wissensmanagement in der Produktentwicklung durch Qualitäts-methodenverbund und Qualitätsmethodenintegration

Band 91
Dr.-Ing. Martin Plass

Beitrag zur Optimierung des Honprozesses durch den Aufbau einer Honprozessregelung

Band 92
Dr.-Ing. Titus Konold

Optimierung der Fünfachsfräsbearbeitung durch eine kennzahlen-unterstützte CAM-Umgebung

Band 93
Dr.-Ing. Jürgen Brath

Unterstützung der Produktionsplanung in der Halbleiterfertigung durch risikoberücksichtigende Betriebskennlinien

Band 94
Dr.-Ing. Dirk Geisinger

Ein Konzept zur marktorientierten Produktentwicklung

Band 95
Dr.-Ing. Marco Lanza

Entwurf der Systemunterstützung des verteilten Engineering mit Axiomatic Design

Band 96
Dr.-Ing. Volker Hüntrup

Untersuchungen zur Mikrostrukturierbarkeit von Stählen durch das Ferti-gungsverfahren Fräsen

Band 97
Dr.-Ing. Frank Reinboth

Interne Stützung zur Genauigkeitsverbesserung in der Inertialmesstechnik: Beitrag zur Senkung der Anforderungen an Inertialsensoren

Band 98
Dr.-Ing. Lutz Trender

Entwicklungsintegrierte Kalkulation von Produktlebenszykluskosten auf Basis der ressourcenorientierten Prozesskostenrechnung

Band 99
Dr.-Ing. Cornelia Kafka

Konzeption und Umsetzung eines Leitfadens zum industriellen Einsatz von Data-Mining

Band 100
Dr.-Ing. Gebhard Selinger

Rechnerunterstützung der informellen Kommunikation in verteilten Unternehmensstrukturen

Band 101
Dr.-Ing. Thomas Windmüller

Verbesserung bestehender Geschäftsprozesse durch eine mitarbeiterorientierte Informationsversorgung

Band 102
Dr.-Ing. Knud Lembke

Theoretische und experimentelle Untersuchung eines bistabilen elektrohydraulischen Linearantriebs

Band 103
Dr.-Ing. Ulrich Thies

Methode zur Unterstützung der variantengerechten Konstruktion von industriell eingesetzten Kleingeräten

Band 104
Dr.-Ing. Andreas Schmälzle

Bewertungssystem für die Generalüberholung von Montageanlagen –Ein Beitrag zur wirtschaftlichen Gestaltung geschlossener Facility- Managment-Systeme im Anlagenbau

Band 105
Dr.-Ing. Thorsten Frank

Vergleichende Untersuchungen schneller elektromechanischer Vorschubachsen mit Kugelgewindetrieb

Band 106
Dr.-Ing. Achim Agostini

Reihenfolgeplanung unter Berücksichtigung von Interaktionen: Beitrag zur ganzheitlichen Strukturierung und Verarbeitung von Interaktionen von Bearbeitungsobjekten

Band 107
Dr.-Ing. Thomas Barrho

Flexible, zeitfenstergesteuerte Auftragseinplanung in segmentierten Fertigungsstrukturen

Band 108
Dr.-Ing. Michael Scharer

Quality Gate-Ansatz mit integriertem Risikomanagement

Band 109
Dr.-Ing. Ulrich Suchy

Entwicklung und Untersuchung eines neuartigen Mischkopfes für das Wasser Abrasivstrahlschneiden

Band 110
Dr.-Ing. Sellal Mussa

Aktive Korrektur von Verlagerungsfehlern in Werkzeugmaschinen

Band 111
Dr.-Ing. Andreas Hühsam

Modellbildung und experimentelle Untersuchung des Wälzschälprozesses

Band 112
Dr.-Ing. Axel Plutowsky

Charakterisierung eines optischen Messsystems und den Bedingungen des Arbeitsraums einer Werkzeugmaschine

Band 113
Dr.-Ing. Robert Landwehr

Konsequent dezentralisierte Steuerung mit Industrial Ethernet und offenen Applikationsprotokollen

Band 114
Dr.-Ing. Christoph Dill

Turbulenzreaktionsprozesse

Band 115
Dr.-Ing. Michael Baumeister

Fabrikplanung im turbulenten Umfeld

Band 116
Dr.-Ing. Christoph Gönnheimer

Konzept zur Verbesserung der Elektromagnetischen Verträglichkeit (EMV) in Produktionssystemen durch intelligente Sensor/Aktor-Anbindung

Band 117
Dr.-Ing. Lutz Demuß

Ein Reifemodell für die Bewertung und Entwicklung von Dienstleistungsorganisationen: Das Service Management Maturity Modell (SMMM)

Band 118
Dr.-Ing. Jörg Söhner

Beitrag zur Simulation zerspanungstechnologischer Vorgänge mit Hilfe der Finite-Element-Methode

Band 119
Dr.-Ing. Judith Elsner

Informationsmanagement für mehrstufige Mikro-Fertigungsprozesse

Band 120
Dr.-Ing. Lijing Xie

Estimation Of Two-dimension Tool Wear Based On Finite Element Method

Band 121
Dr.-Ing. Ansgar Blessing

Geometrischer Entwurf mikromechatronischer Systeme

Band 122
Dr.-Ing. Rainer Ebner

Steigerung der Effizienz mehrachsiger Fräsprozesse durch neue Planungsmethoden mit hoher Benutzerunterstützung

Band 123
Dr.-Ing. Silja Klinkel

Multikriterielle Feinplanung in teilautonomen Produktionsbereichen – Ein Beitrag zur produkt- und prozessorientierten Planung und Steuerung

Band 124
Dr.-Ing. Wolfgang Neithardt

Methodik zur Simulation und Optimierung von Werkzeugmaschinen in der Konzept- und Entwurfsphase auf Basis der Mehrkörpersimulation

Band 125
Dr.-Ing. Andreas Mehr

Hartfeinbearbeitung von Verzahnungen mit kristallinen diamantbeschichteten Werkzeugen beim Fertigungsverfahren Wälzstoßen

Band 126
Dr.-Ing. Martin Gutmann

Entwicklung einer methodischen Vorgehensweise zur Diagnose von hydraulischen Produktionsmaschinen

Band 127
Dr.-Ing. Gisela Lanza

Simulative Anlaufunterstützung auf Basis der Qualitätsfähigkeiten von Produktionsprozessen

Band 128
Dr.-Ing. Ulf Dambacher

Kugelgewindetrieb mit hohem Druckwinkel

Band 129
Dr.-Ing. Carsten Buchholz

Systematische Konzeption und Aufbau einer automatisierten Produktionszelle für pulverspritzgegossene Mikrobauteile

Band 130
Dr.-Ing. Heiner Lang

Trocken-Räumen mit hohen Schnittgeschwindigkeiten

Band 131
Dr.-Ing. Daniel Nesges

Prognose operationeller Verfügbarkeiten von Werkzeugmaschinen unter Berücksichtigung von Serviceleistungen

Im Shaker Verlag erschienene Bände:

Band 132
Dr.-Ing. Andreas Bechle

Beitrag zur prozesssicheren Bearbeitung beim Hochleistungs-fertigungsverfahren Wälzschälen

Band 133
Dr.-Ing. Markus Herm

Konfiguration globaler Wertschöpfungsnetzwerke auf Basis von Business Capabilities

Band 134
Dr.-Ing. Hanno Tritschler

Werkzeug- und Zerspanprozessoptimierung beim Hartfräsen von Mikrostrukturen in Stahl

Band 135
Dr.-Ing. Christian Munzinger

Adaptronische Strebe zur Steifigkeitssteigerung von Werkzeugmaschinen

Band 136
Dr.-Ing. Andreas Stepping

Fabrikplanung im Umfeld von Wertschöpfungsnetzwerken und ganzheitlichen Produktionssystemen

Band 137
Dr.-Ing. Martin Dyck

Beitrag zur Analyse thermische bedingter Werkstückdeformationen in Trockenbearbeitungsprozessen

Band 138
Dr.-Ing. Siegfried Schmalzried

Dreidimensionales optisches Messsystem für eine effizientere geometrische Maschinenbeurteilung

Band 139
Dr.-Ing. Marc Wawerla

Risikomanagement von Garantieleistungen

Band 140
Dr.-Ing. Ivesa Buchholz

Strategien zur Qualitätssicherung mikromechanischer Bauteile mittels multisensorieller Koordinatenmesstechnik

Band 141
Dr.-Ing. Jan Kotschenreuther

Empirische Erweiterung von Modellen der Makrozerspanung auf den Bereich der Mikrobearbeitung

Band 142
Dr.-Ing. Andreas Knödel

Adaptronische hydrostatische Drucktascheneinheit

Band 143
Dr.-Ing. Gregor Stengel

Fliegendes Abtrennen räumlich gekrümmter Strangpressprofile mittels Industrierobotern

Band 144
Dr.-Ing. Udo Weismann

Lebenszyklusorientiertes interorganisationelles Anlagencontrolling

Band 145
Dr.-Ing. Rüdiger Pabst

Mathematische Modellierung der Wärmestromdichte zur Simulation des thermischen Bauteilverhaltens bei der Trockenbearbeitung

Band 146
Dr.-Ing. Jan Wieser

Intelligente Instandhaltung zur Verfügbarkeitssteigerung von Werkzeugmaschinen

Band 147
Dr.-Ing. Sebastian Haupt

Effiziente und kostenoptimale Herstellung von Mikrostrukturen durch eine Verfahrenskombination von Bahnerosion und Laserablation

Band 148
Dr.-Ing. Matthias Schlipf

Statistische Prozessregelung von Fertigungs- und Messprozess zur Erreichung einer variabilitätsarmen Produktion mikromechanischer Bauteile

Band 149
Dr.-Ing. Jan Philipp Schmidt-Ewig

Methodische Erarbeitung und Umsetzung eines neuartigen Maschinenkonzeptes zur produktflexiblen Bearbeitung räumlich gekrümmter Strangpressprofile

Band 150
Dr.-Ing. Thomas Ender

Prognose von Personalbedarfen im Produktionsanlauf unter Berücksichtigung dynamischer Planungsgrößen

Band 151
Dr.-Ing. Kathrin Peter

Bewertung und Optimierung der Effektivität von Lean Methoden in der Kleinserienproduktion

Band 152
Dr.-Ing. Matthias Schopp

Sensorbasierte Zustandsdiagnose und -prognose von Kugelgewindetrieben

Band 153
Dr.-Ing. Martin Kipfmüller

Aufwandsoptimierte Simulation von Werkzeugmaschinen

Band 154
Dr.-Ing. Carsten Schmidt

Development of a database to consider multi wear mechanisms within chip forming simulation

Band 155
Dr.-Ing. Stephan Niggeschmidt

Ausfallgerechte Ersatzteilbereitstellung im Maschinen- und Anlagenbau mittels lastabhängiger Lebensdauerprognose

Band 156
Dr.-Ing. Jochen Conrad Peters

Bewertung des Einflusses von Formabweichungen in der Mikro-Koordinatenmesstechnik

Band 157
Dr.-Ing. Jörg Ude

Entscheidungsunterstützung für die Konfiguration globaler Wertschöpfungsnetzwerke

Band 158
Dr.-Ing. Stefan Weiler

Strategien zur wirtschaftlichen Gestaltung der globalen Beschaffung

Band 159
Dr.-Ing. Jan Rühl

Monetäre Flexibilitäts- und Risikobewertung

Band 160
Dr.-Ing. Daniel Ruch

Positions- und Konturerfassung räumlich gekrümmter Profile auf Basis bauteilimmanenter Markierungen

Band 161
Dr.-Ing. Manuel Tröndle

Flexible Zuführung von Mikrobauteilen mit piezoelektrischen Schwingförderern

Band 162
Dr.-Ing. Benjamin Viering

Mikroverzahnungsnormal

Band 163
Dr.-Ing. Chris Becke

Prozesskraftrichtungsangepasste Frässtrategien zur schädigungsarmen Bohrungsbearbeitung an faserverstärkten Kunststoffen

Band 164
Dr.-Ing. Patrick Werner

Dynamische Optimierung und Unsicherheitsbewertung der lastabhängigen präventiven Instandhaltung von Maschinenkomponenten

Band 165
Dr.-Ing. Martin Weis

Kompensation systematischer Fehler bei Werkzeugmaschinen durch self-sensing Aktoren

Band 166
Dr.-Ing. Markus Schneider

Kompensation von Konturabweichungen bei gerundeten Strangpressprofilen durch robotergestützte Führungswerkzeuge

Band 167
Dr.-Ing. Ester M. R. Ruprecht

Prozesskette zur Herstellung schichtbasierter Systeme mit integrierten Kavitäten

Band 168
Dr.-Ing. Alexander Broos

Simulationsgestützte Ermittlung der Komponentenbelastung für die Lebensdauerprognose an Werkzeugmaschinen

Band 169
Dr.-Ing. Frederik Zanger

Segmentspanbildung, Werkzeugverschleiß, Randschichtzustand und Bauteileigenschaften: Numerische Analysen zur Optimierung des Zerspanungsprozesses am Beispiel von Ti-6Al-4V

Band 170
Dr.-Ing. Benjamin Behmann

Servicefähigkeit

Band 171
Dr.-Ing. Annabel Gabriele Jondral

Simulationsgestützte Optimierung und Wirtschaftlichkeitsbewertung des Lean-Methodeneinsatzes

Band 172
Dr.-Ing. Christoph Ruhs

Automatisierte Prozessabfolge zur qualitätssicheren Herstellung von Kavitäten mittels Mikrobahnerosion

Band 173
Dr.-Ing. Steven Peters

Markoffsche Entscheidungsprozesse zur Kapazitäts- und Investitionsplanung von Produktionssystemen

Band 174
Dr.-Ing. Christoph Kühlewein

Untersuchung und Optimierung des Wälzschälverfahrens mit Hilfe von 3D-FEM-Simulation – 3D-FEM Kinematik- und Spanbildungssimulation

Band 175
Dr.-Ing. Adam-Mwanga Dieckmann

Auslegung und Fertigungsprozessgestaltung sintergefügter Verbindungen für µMIM-Bauteile

Band 176
Dr.-Ing. Heiko Hennrich

Aufbau eines kombinierten belastungs- und zustandsorientierten Diagnose- und Prognosesystems für Kugelgewindetriebe

Band 177
Dr.-Ing. Stefan Herder

Piezoelektrischer Self-Sensing-Aktor zur Vorspannungsregelung in adaptronischen Kugelgewindetrieben

Band 178
Dr.-Ing. Alexander Ochs

Ultraschall-Strömungsgreifer für die Handhabung textiler Halbzeuge bei der automatisierten Fertigung von RTM-Bauteilen

Band 179
Dr.-Ing. Jürgen Michna

Numerische und experimentelle Untersuchung zerspanungsbedingter Gefügeumwandlungen und Modellierung des thermo-mechanischen Lastkollektivs beim Bohren von 42CrMo4

Band 180
Dr.-Ing. Jörg Elser

Vorrichtungsfreie räumliche Anordnung von Fügepartnern auf Basis von Bauteilmarkierungen

Band 181
Dr.-Ing. Katharina Klimscha

Einfluss des Fügespalts auf die erreichbare Verbindungsqualität beim Sinterfügen

Band 182
Dr.-Ing. Patricia Weber

Steigerung der Prozesswiederholbarkeit mittels Analyse akustischer Emissionen bei der Mikrolaserablation mit UV-Pikosekundenlasern

Band 183
Dr.-Ing. Jochen Schädel

Automatisiertes Fügen von Tragprofilen mittels Faserwickeln

Band 184
Dr.-Ing. Martin Krauße

Aufwandsoptimierte Simulation von Produktionsanlagen durch Vergrößerung der Geltungsbereiche von Teilmodellen

Band 185
Dr.-Ing. Raphael Moser

Strategische Planung globaler Produktionsnetzwerke
Bestimmung von Wandlungsbedarf und Wandlungszeitpunkt mittels multikriterieller Optimierung

Band 186
Dr.-Ing. Martin Otter

Methode zur Kompensation fertigungsbedingter Gestaltabweichungen für die Montage von Aluminium Space-Frame-Strukturen

Band 187
Dr.-Ing. Urs Leberle

Produktive und flexible Gleitförderung kleiner Bauteile auf phasenflexiblen Schwingförderern mit piezoelektrischen 2D-Antriebselementen

Band 188
Dr.-Ing. Johannes Book

Modellierung und Bewertung von Qualitätsmanagementstrategien in globalen Wertschöpfungsnetzwerken

Band 189
Dr.-Ing. Florian Ambrosy

Optimierung von Zerspanungsprozessen zur prozesssicheren Fertigung nanokristalliner Randschichten am Beispiel von 42CrMo4

Band 190
Dr.-Ing. Adrian Kölmel

Integrierte Messtechnik für Prozessketten unreifer Technologien am Beispiel der Batterieproduktion für Elektrofahrzeuge

Band 191
Dr.-Ing. Henning Wagner

Featurebasierte Technologieplanung zum Preforming von textilen Halbzeugen

Band 192
Dr.-Ing. Johannes Gebhardt

Strukturoptimierung von in FVK eingebetteten metallischen Lasteinleitungselementen

Band 193
Dr.-Ing. Jörg Bauer

Hochintegriertes hydraulisches Vorschubsystem für die Bearbeitung kleiner Werkstücke mit hohen Fertigungsanforderungen

Band 194
Dr.-Ing. Nicole Stricker

Robustheit verketteter Produktionssysteme
Robustheitsevaluation und Selektion des Kennzahlensystems der Robustheit

Band 195
Dr.-Ing. Anna Sauer

Konfiguration von Montagelinien unreifer Produkttechnologien am Beispiel der Batteriemontage für Elektrofahrzeuge

Band 196
Dr.-Ing. Florian Sell-Le Blanc

Prozessmodell für das Linearwickeln unrunder Zahnspulen
Ein Beitrag zur orthozyklischen Spulenwickeltechnik

Band 197
Dr.-Ing. Frederic Förster

Geregeltes Handhabungssystem zum zuverlässigen und energieeffizienten Handling textiler Kohlenstofffaserzuschnitte

Band 198
Dr.-Ing. Nikolay Boev

Numerische Beschreibung von Wechselwirkungen zwischen Zerspanprozess und Maschine am Beispiel Räumen

Band 199
Dr.-Ing. Sebastian Greinacher

Simulationsgestützte Mehrzieloptimierung schlanker und ressourceneffizienter Produktionssysteme

Band 200
Dr.-Ing. Benjamin Häfner

Lebensdauerprognose in Abhängigkeit der Fertigungsabweichungen bei Mikroverzahnungen

Band 201
Dr.-Ing. Stefan Klotz

Dynamische Parameteranpassung bei der Bohrungsherstellung in faserverstärkten Kunststoffen unter zusätzlicher Berücksichtigung der Einspannsituation

Band 202
Dr.-Ing. Johannes Stoll

Bewertung konkurrierender Fertigungsfolgen mittels Kostensimulation und stochastischer Mehrzieloptimierung
Anwendung am Beispiel der Blechpaketfertigung für automobile Elektromotoren

Band 203
Dr.-Ing. Simon-Frederik Koch

Fügen von Metall-Faserverbund-Hybridwellen im Schleuderverfahren
ein Beitrag zur fertigungsgerechten intrinsischen Hybridisierung

Band 204
Dr.-Ing. Julius Ficht

Numerische Untersuchung der Eigenspannungsentwicklung für sequenzielle Zerspanungsprozesse

Band 205
Dr.-Ing. Manuel Baumeister

Automatisierte Fertigung von Einzelblattstapeln in der Lithium-Ionen-Zellproduktion

Band 206
Dr.-Ing. Daniel Bertsch

Optimierung der Werkzeug- und Prozessauslegung für das Wälzschälen von Innenverzahnungen

Band 207
Dr.-Ing. Kyle James Kippenbrock

Deconvolution of Industrial Measurement and Manufacturing Processes for Improved Process Capability Assessments

Band 208
Dr.-Ing. Farboud Bejnoud

Experimentelle Prozesskettenbetrachtung für Räumbauteile am Beispiel einer einsatzgehärteten PKW-Schiebemuffe

Band 209
Dr.-Ing. Steffen Dosch

Herstellungsübergreifende Informationsübertragung zur effizienten Produktion von Werkzeugmaschinen am Beispiel von Kugelgewindetrieben

Band 210
Dr.-Ing. Emanuel Moser

Migrationsplanung globaler Produktionsnetzwerke
Bestimmung robuster Migrationspfade und risiko-effizienter Wandlungsbefähiger

Band 211
Dr.-Ing. Jan Hochdörffer

Integrierte Produktallokationsstrategie und Konfigurationssequenz in globalen Produktionsnetzwerken

Band 212
Dr.-Ing. Tobias Arndt

Bewertung und Steigerung der Prozessqualität in globalen Produktionsnetzwerken

Band 213
Dr.-Ing. Manuel Peter

Unwuchtminimale Montage von Permanentmagnetrotoren durch modellbasierte Online-Optimierung

Band 214
Dr.-Ing. Robin Kopf

Kostenorientierte Planung von Fertigungsfolgen additiver Technologien

Band 215
Dr.-Ing. Harald Meier

Einfluss des Räumens auf den Bauteilzustand in der Prozesskette Weichbearbeitung – Wärmebehandllung – Hartbearbeitung

Band 216
Dr.-Ing. Daniel Brabandt

Qualitätssicherung von textilen Kohlenstofffaser-Preforms mittels optischer Messtechnik

Band 217
Dr.-Ing. Alexandra Schabunow

Einstellung von Aufnahmeparametern mittels projektionsbasierter Qualitätskenngrößen in der industriellen Röntgen-Computertomographie

Band 218
Dr.-Ing. Jens Bürgin

Robuste Auftragsplanung in Produktionsnetzwerken
Mittelfristige Planung der variantenreichen Serienproduktion unter Unsicherheit der Kundenauftragskonfigurationen

Band 219
Dr.-Ing. Michael Gerstenmeyer

Entwicklung und Analyse eines mechanischen Oberflächenbehandlungsverfahrens unter Verwendung des Zerspanungswerkzeuges

Band 220
Dr.-Ing. Jacques Burtscher

Erhöhung der Bearbeitungsstabilität von Werkzeugmaschinen durch semi-passive masseneinstellbare Dämpfungssysteme

Band 221
Dr.-Ing. Dietrich Berger

Qualitätssicherung von textilen Kohlenstofffaser-Preforms mittels prozessintegrierter Wirbelstromsensor-Arrays

Band 222
Dr.-Ing. Fabian Johannes Ballier

Systematic gripper arrangement for a handling device in lightweight production processes

Band 223
Dr.-Ing. Marielouise Schäferling, geb. Zaiß

Development of a Data Fusion-Based Multi-Sensor System for Hybrid Sheet Molding Compound

Band 224
Dr.-Ing. Quirin Spiller

Additive Herstellung von Metallbauteilen mit dem ARBURG Kunststoff-Freiformen

Band 225
Dr.-Ing. Andreas Spohrer

Steigerung der Ressourceneffizienz und Verfügbarkeit von Kugelgewindetrieben durch adaptive Schmierung

Band 226
Dr.-Ing. Johannes Fisel

Veränderungsfähigkeit getakteter Fließmontagesysteme
Planung der Fließbandabstimmung am Beispiel der Automobilmontage

Band 227
Dr.-Ing. Patrick Bollig

Numerische Entwicklung von Strategien zur Kompensation thermisch bedingter Verzüge beim Bohren von 42CrMo4

Band 228
Dr.-Ing. Ramona Pfeiffer, geb. Singer

Untersuchung der prozessbestimmenden Größen für die anforderungsgerechte Gestaltung von Pouchzellen-Verpackungen

Band 229
Dr.-Ing. Florian Baumann

Additive Fertigung von endlosfaserverstärkten Kunststoffen mit dem ARBURG Kunststoff-Freiform Verfahren

Band 230
Dr.-Ing. Tom Stähr

Methodik zur Planung und Konfigurationsauswahl skalierbarer Montagesysteme – Ein Beitrag zur skalierbaren Automatisierung

Band 231
Dr.-Ing. Jan Schwennen

Einbringung und Gestaltung von Lasteinleitungsstrukturen für im RTM-Verfahren hergestellte FVK-Sandwichbauteile

Band 232
Dr.-Ing. Sven Coutandin

Prozessstrategien für das automatisierte Preforming von bebinderten textilen Halbzeugen mit einem segmentierten Werkzeugsystem

Band 233
Dr.-Ing. Christoph Liebrecht

Entscheidungsunterstützung für den Industrie 4.0-Methodeneinsatz
Strukturierung, Bewertung und Ableitung von Implementierungsreihenfolgen

Band 234
Dr.-Ing. Stefan Treber

Transparenzsteigerung in Produktionsnetzwerken
Verbesserung des Störungsmanagements durch verstärkten Informationsaustausch

Band 235
Dr.-Ing. Marius Dackweiler

Modellierung des Fügewickelprozesses zur Herstellung von leichten Fachwerkstrukturen

Band 236
Dr.-Ing. Fabio Echsler Minguillon

Prädiktiv-reaktives Scheduling zur Steigerung der Robustheit in der Matrix-Produktion

Band 237
Dr.-Ing. Sebastian Haag

Entwicklung eines Verfahrensablaufes zur Herstellung von Batteriezellstapeln mit großformatigem, rechteckigem Stapelformat und kontinuierlichen Materialbahnen

Band 238
Dr.-Ing. Raphael Wagner

Strategien zur funktionsorientierten Qualitätsregelung in der Serienproduktion

Band 239
Dr.-Ing. Christopher Ehrmann

Ausfallfrüherkennung von Ritzel-Zahnstangen- Trieben mittels Acoustic Emission

Band 240
Dr.-Ing. Janna Hofmann

Prozessmodellierung des Fünf-Achs-Nadelwickelns zur Implementierung einer trajektoriebasierten Drahtzugkraftregelung

Band 241
Dr.-Ing. Andreas Kuhnle

Adaptive Order Dispatching based on Reinforcement Learning
Application in a Complex Job Shop in the Semiconductor Industry

Band 242
Dr.-Ing. Andreas Greiber

Fertigung optimierter technischer Oberflächen durch eine Verfahrenskombination aus Fliehkraft-Tauchgleitschleifen und Laserablation
Prozesseinflüsse und Prozessauslegung

Band 243
Dr.-Ing. Jan Niclas Eschner

Entwicklung einer akustischen Prozessüberwachung zur Porenbestimmung im Laserstrahlschmelzen

Band 244
Dr.-Ing. Sven Roth

Schädigungsfreie Anbindung von hybriden FVK/Metall-Bauteilen an metallische Tragstrukturen durch Widerstandspunktschweißen

Band 245
Dr.-Ing. Sina Kathrin Peukert

Robustheitssteigerung in Produktionsnetzwerken mithilfe eines integrierten Störungsmanagements

Band 246
Dr.-Ing. Alexander Jacob

Hochiterative Technologieplanung
Rekursive Optimierung produkt- und fertigungsbezogener
Freiheitsgrade am Beispiel der hybrid-additiven Fertigung

Band 247
Dr.-Ing. Patrick Moll

Ressourceneffiziente Herstellung von Langfaser-Preforms im Faserblasverfahren

Band 248
Dr.-Ing. Eric Thore Segebade

Erhöhung der Verschleißbeständigkeit von Bauteilen aus Ti-6Al-4V mittels simulationsgestützer Zerspanung und mechanischer Mikrotexturierung

Band 249
Dr.-Ing. Shun Yang

Regionalized implementation strategy of smart automation within assembly systems in China

Band 250
Dr.-Ing. Constantin Carl Hofmann

Vorausschauende und reaktive Mehrzieloptimierung für die Produktionssteuerung einer Matrixproduktion

Band 251
Dr.-Ing. Paul Ruhland

Prozesskette zur Herstellung von hybriden Faser-Metall-Preforms
Modellbildung und Optimierung des Binderauftrags und
der Drapierung für stabförmige Bauteile

Band 252
Dr.-Ing. Leonard Schild

Erzeugung und Verwendung von Anwendungswissen in der industriellen Computertomographie

Band 253
Dr.-Ing. Benedikt Klee

Analyse von Phaseninformationen in Videodaten zur Identifikation von Schwingungen in Werkzeugmaschinen

Band 254
Dr.-Ing. Bruno Vargas

Wälzschälen mit kleinen Achskreuzwinkeln
Prozessgrenzen und Umsetzbarkeit

Band 255
Dr.-Ing. Lucas Bretz

Function-oriented in-line quality assurance of hybrid sheet molding compound

Band 256
Dr.-Ing. Bastian Rothaupt

Dämpfung von Bauteilschwingungen durch einstellbare Werkstückdirektspannung mit Hydrodehnspanntechnik

Band 257
Dr.-Ing. Daniel Kupzik

Robotic Swing Folding of three-dimensional UD-tape-based Reinforcement Structures

Band 258
Dr.-Ing. Bastian Verhaelen

(De-)Zentralisierung von Entscheidungen in globalen Produktionsnetzwerken
Strategie- und komplexitätsorientierte Gestaltung der Entscheidungsautonomie

Band 259
Dr.-Ing. Hannes Wilhelm Weinmann

Integration des Vereinzelungs- und Stapelbildungsprozesses in ein flexibel und kontinuierlich arbeitendes Anlagenmodul für die Li-Ionen Batteriezellfertigung

Band 260
Dr.-Ing. Florian Stamer

Dynamische Lieferzeit-Preisgestaltung in variantenreicher Produktion
Ein adaptiver Ansatz mithilfe von Reinforcement Learning

Band 261
Dr.-Ing. Patrick Neuenfeldt

Modellbildung des Tauchgleitschleifens zur Abtrag- und Topografievorhersage an komplexen Geometrien

Band 262
Dr.-Ing. Boris Matuschka

Energieeffizienz in Prozessketten: Analyse und Optimierung von Energieflüssen bei der Herstellung eines PKW-Getriebebauteils aus 16MnCr5

Band 263
Dr.-Ing. Tobias Schlagenhauf

Bildbasierte Quantifizierung und Prognose des Verschleißes an Kugelgewindetriebspindeln
Ein Beitrag zur Zustandsüberwachung von Kugelgewindetrieben mittels Methoden des maschinellen Lernens

Band 264
Dr.-Ing. Benedict Stampfer

Entwicklung eines multimodalen Prozessmodells zur Oberflächenkonditionierung beim Außenlängsdrehen von 42CrMo4

Band 265
Dr.-Ing. Carmen Maria Krahe

KI-gestützte produktionsgerechte Produktentwicklung
Automatisierte Wissensextraktion aus vorhandenen Produktgenerationen

Band 266
Dr.-Ing. Markus Netzer

Intelligente Anomalieerkennung für hochflexible Produktionsmaschinen
Prozessüberwachung in der Brownfield Produktion

Band 267
Dr.-Ing. Simon Raphael Merz

Analyse der Kinematik und Kinetik von Planetenwälzgewindetrieben

Band 268
Dr.-Ing. Rainer Maria Silbernagel

Funktionsorientierte Qualitätsregelung in Produktionsnetzwerken
Qualitätsmanagement in der Produktion hochpräziser Produkte durch netzwerkweite Datenintegration

Band 269
Dr.-Ing. Jonas Nieschlag

Gestaltung und Prozessanalyse für im Schleuderverfahren hergestellte FKV-Metall-Hohlstrukturen

Band 270
Dr.-Ing. Lukas Matthias Weiser

In-Process Porositätserkennung für den PBF-LB/M-Prozess

Band 271
Dr.-Ing. Leonard Vincent Overbeck

Digital Twins of production systems
Automated validation and update of material flow simulation models with real data

wbk Institut für Produktionstechnik
Karlsruher Institut für Technologie (KIT)

Zum Buch

To achieve good economic efficiency and sustainability, production systems must be operated at a high level of productivity over long periods. This poses great challenges for manufacturing companies, especially in times of increased volatility, caused, amongst others, by the technological transformation in the mobility sector, as well as political and social change, which lead to constantly evolving requirements on the production system. Because the frequency of necessary adaptation decisions and subsequent optimization measures is increasing, the need for evaluation capabilities of scenarios and possible system configurations is growing. A widely applicable, powerful tool for this purpose is material flow simulation, but its use is currently limited by its time-consuming manual creation and its limited, project-based usage. A long-term, lifecycle accompanying use is currently hindered by the simulation model's labor-intensive maintenance, i.e. the model's manual adaptation in case of changes in the real system.
This thesis aims to develop and implement a concept including the necessary methods to automate the simulation model's maintenance and adaptation to reality and improve the model's accuracy. For this purpose, digital data from the real system are used, which are increasingly available due to trends such as Industry 4.0 and digitalization in general. The pursued vision of this work is a Digital Twin of the production system, which represents a realistic image of the system in the long term through the data-based comparison with reality and its adaptation to reality. This Digital Twin can be used for the realistic evaluation of scenarios, actions, and improvement measures. Therefore, an overall concept and mechanisms for automatic validation and updating of the model were developed. Among other things, the focus was on the development of algorithms for the detection of changes in the structure and processes in the production system, as well as on the study of the influence of the available data on the achievable quality of the Digital Twin.
The developed components could be successfully applied to a real industrial use case at the Robert Bosch GmbH where it lead to a high accuracy Digital Twin, which was successfully used for production planning and improvement. The potential of localization data for the creation of Digital Twins of production systems could be shown in the laboratory environment of the learning factory at the wbk Institute of Production Science.

ISBN 978-3-8440-9242-4

9 783844 092424

SHAKER
VERLAG