19th IMEKO TC10 Conference
"MACRO meets NANO in Measurement for Diagnostics, Optimization and Control"
Delft, The Netherlands, September 21–22, 2023

Improving the Planning Quality in Practice with Artificial Intelligence

Leonhard Czarnetzki[1], David Karnok[2], Johannes Breitschopf[3] Matthias Karner[4], Milot Gashi[5], Thomas Mambrini[6], Catherine Laflamme[6], Viola Gallina[8]

[1] *Fraunhofer Austria Research GmbH, Wattens, Austria, leonhard.czarnetzki@fraunhofer.at*
[2] *SZTAKI, Budapest, Hungary, david.karnok@sztaki.hu*
[3] *Fraunhofer Austria Research GmbH, Vienna, Austria, johannes.breitschopf@fraunhofer.at*
[4] *voestalpine BÖHLER Bleche GmbH, Mürzzuschlag, Austria, matthias.karner@voestalpine.com*
[5] *AT&S Austria Technologie & Systemtechnik, Leoben, Austria, m.gashi@ats.net*
[6] *AT&S Austria Technologie & Systemtechnik, Leoben, Austria, t.mambrini@ats.net*
[7] *Fraunhofer Austria Research GmbH, Vienna, Austria, catherine.laflamme@fraunhofer.at*
[8] *Fraunhofer Austria Research GmbH, Vienna, Austria, viola.gallina@fraunhofer.at*

Abstract –
Accurate production planning is economic interest of manufacturing companies. Reducing the work-in-progress levels, the lead time or control efforts with the simultaneous increase of utilization and adherence to schedule might lead to instantaneous cost reduction and to increased competitiveness on long-term. In the era of digitization various artificial intelligence-based methods have been investigated by the scientific community to improve these key performance indicators. In this paper the results of a joint research project dealing with planning quality improvement with the help of Machine Learning (ML) are summarized. The results of two use case studies investigating the application and suitability of different planning approaches in the semiconductor and steel industries are presented and considerations regarding the practical application of ML assisted planning approaches are discussed.

Keywords – Production scheduling; Production planning and control; Master data; Dynamic data; Optimization; Industry, Innovation and Infrastructure.*

I. INTRODUCTION

Production Planning and Control (PPC) coordinates all relevant activities along the order processing chain to ensure that the predefined logistic KPIs, such as delivery times, inventory levels or capacity utilization, are in the acceptable range [1]. As achieving performance targets can have short and long-term economic consequences for manufacturing companies, reliable and accurate production planning is vital. However, unforeseen events, such as machine failures or delays in material delivery, insufficient underlying planning data or inappropriate planning and control systems itself can cause deviations between a production plan and its execution. A common countermeasure is the increased use of buffers, which contrarily not only increases coordination and control efforts but might lead to negative effects, such as higher inventory levels, higher throughput time or lower capacity utilization [2]. Instead, increasing the accuracy and reliability of the production plan and its execution can reduce the number and impact of deviations and therefore help to achieve the desired performance targets. Conveniently, the advancement of digitalization in companies and the increasing use of IoT-devices in manufacturing makes more and more data on the relevant processes available, opening the door for data-driven planning methods and the use of new technologies, such as Machine Learning (ML). However, while promising solutions to isolated problems exist in the literature, several open questions still remain, e.g. how to best integrate them into a holistic approach to PPC for cyber-physical production systems [3]. The results presented in this paper are on the one hand results of the national research project MLinPPC and on the other hand the continuation of previous works – started before the MLinPPC project and to be continued after the end of this project – in the fields of cyber-physical production systems and situation aware PPC. The remainder of the paper is organized as follows: in chapter II we review previous work of the authors in the field of PPC assisted by Machine Learning (ML) and other related literature, in chapter III we show the results of an industrial use case study for two different planning approaches, that systematically and iteratively improve the planning quality (PQ) of a production plan with each planning cycle based on ML methods. In chapter IV, we discuss considerations regarding the practical application of ML-based planning approaches. Finally, in chapter V, we summarize our results and provide an outlook.

II. STATE-OF-THE-ART

In previous works, the authors showed that production planning based on dynamic data – generated with various machine learning (ML) algorithms – potentially outperforms classical, static and analytical approaches (e.g., work in progress prediction with LSTM [4], lead time prediction with tree-based methods and regression [5]). The conducted research involved different ways of work from the beginning of simulation [6], through digital twin [6, 7] to the end of pure historical data analysis and usage [5]. The authors made a first try to define the planning quality (PQ) and differentiate it from robustness [8]. A common problem for data and ML based planning methods is low data quality. The main causes for low data quality are typically recent (within the past 10 years) digitalization of a company and due to the second nature data collection is treated at these companies. Improving upon these aspects usually involve assigning a cost to having low quality data and better operator training [9]. When considering data quality and quantity together, there is usually a trade-off between the two in practice: large amounts of low-quality data, small amounts of high-quality data and anything in between. Doing machine learning on its own at the two extremes can be troublesome. The first extreme is handled by extensive data preparations and the latter is handled by introducing simulation. (Note that the preparation may move the needle into the latter direction, requiring the ML to be supplemented by simulation anyway). The simulation models in these cases are somewhat bounded in detail and highly depend on the underlying company, process and properties one wishes to learn [10]. Digitalization does help, though often indirectly, to improve competitiveness via its better transparency of the operations data at a company. However, the scale of the company and the digitalization amount to be supplemented by more digitalization is not a linear relationship, nor is the cost associated. SMEs have better chance to become digitalized whereas large and established companies have more stumbling blocks (from management to the shop floor) counteracting such changes [11]. The resistance to (the expansion of) digitalization at a company is often not technological in nature, but human reluctance to "change what's working fine" from one's perspective. Depending on the company's general properties (i.e., archetype), education, simulation and demonstration of what the current and future data collection may enable, along with other incentives, can ease the transition considerably [12].

III. PLANNING APPROACHES FOR IMPROVING THE PLANNING QUALITY

The main idea of the MLinPPC project was to test approaches that systematically increase the PQ with the help of ML. For this purpose, two approaches were foreseen: the evolutionary and the function-based approach. While in case of the previous one the PQ is expected to increase iteratively with the *adjustment of the planning data*, the later one has an inherently high PQ and includes the *update of the production schedule* as well (for more information please see [3], [13]). In the following two subchapters an industrial use case study for each approach is presented.

A. Evolutionary approach

The first use case study applies the evolutionary approach (see Figure 1) to improve PPC by using ML approaches. As shown in Figure 1, this approach is defined by improving the PQ iteratively, whenever the need for it is recognized. A visualization tool is used to evaluate the deviations from the production plan and therefore to determine, whether a new improvement cycle must be initiated. The study is based in the semiconductor industry and focuses on a specific process step within the production. In this process step deviations from the original production plan are especially high. Therefore, ML based forecasting is used to predict whether a lot will arrive at a particular workstation within a predefined number of shifts.

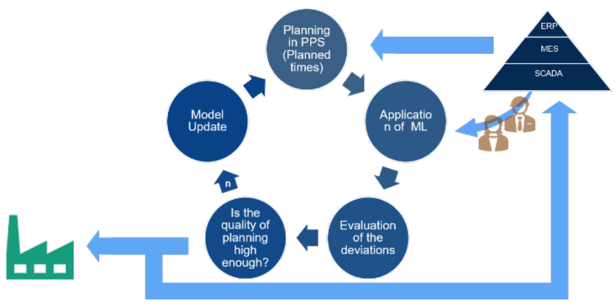

Figure 1 An overview of the evolutionary approach use case

As a basis for training and evaluation of the ML based planning method, we use confirmation data of the workstations under investigation. This data describes whether the processing of a specific lot is completed at a specific workstation. We consider state-of-the-art ensemble learning approaches, such as lightGBM and further plan to conduct a comparison of various other ML algorithms. Finally, to evaluate the performance of the models, we consider various metrics such as accuracy (acc), F1-score, precision, and recall. The choice of this metrics helps us to take into account the challenge of imbalanced data, that is expected to be present within the historical as well as the future data. In Figure 2, the evaluation results of the lightGBM model for a 48-hour prediction horizon are shown. In Figure 2-a, the prediction results (accuracy) from different iterations are shown as box plots, where different colours represent different metrics. Furthermore, Figure 2-b shows the model evaluation of each iteration explicitly. There it can be seen that with each iteration the performance of the ML model increases. Moreover, the performance strongly correlates with the amount of data used for training the model. In general, the results show that the explored problem is predictable by ML although for the application of the model in the production process better results are required. Therefore, the development of the prediction model is still work in progress. As a next step we plan to analyse what is the optimal trade-off between the training data- size and the accuracy which is important when aiming to develop an evolutionary approach. Furthermore, we plan to investigate on how we can detect special time-based events in the data, such as Christmas, that highly influence the production. In this regard, we will consider outlier and anomaly detection approaches.

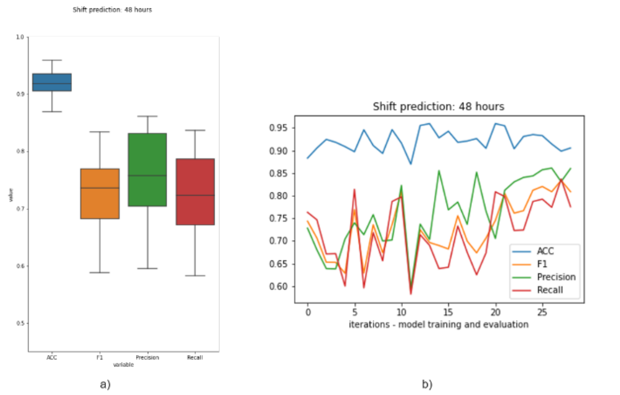

Figure 2 Evaluation results using lightGBM model and a prediction horizon of 48 hours.

As a visualization tool, a dashboard was developed using the open-source framework Dash, implemented in Python. The raw data is first pre-processed in an independent python script, with the output data saved locally as input for the dashboard. The dashboard includes four windows for the following topics: Lead time, Gantt chart, WIP and the accumulation. In the lead time window, a histogram of all lead times (calculated per job) is given, with a color coding based on different categorical properties, such as the plant ID and the product group. The Gantt chart displays a given selection of jobs (selected from the entire job list) and is broken down into the different stations at which the product is produced. As jobs are added in a color-coded way, this visualization allows for the quick identification of utilization of stations over time. The WIP visualization is the sum of products which are finished at a particular workstation per day. By summing over all workstations, and comparing individual workstations with another, one can easily see the overall capacity of production, and identify bottlenecks and variations in production. Finally, the accumulation plot tracks the progress of production at a given workstation over time. With finished products being added with a negative sign, the influx and outflux of products at a workstation can be observed through the course of the day. Figure 3 shows a screenshot of the visualization tool.

Figure 3 An example of the AxoLOTI showing ML model evaluation over the iteration.

B. Function based approach

The second use case of the study applies the function-based approach to improve the PPS of a duo rolling mill for the production of high-quality steel and titanium plates. The production process can be described as follows. The plates are heated to a specified temperature before rolling can begin (the rolling temperature varies depending on the product). Then, a few rolling passes take place in which the thickness is reduced and the length is increased. During rolling the material cools down; once a minimum temperature is reached, the plates are again reheated in the furnace to rolling temperature. This cycle (one cycle is called a heat) is repeated until the target thickness is reached.

As the production process and its parameters vary depending on the alloys and dimensions, the following restrictions must be considered: the furnace temperatures, heat requirements, plate dimensions for the optimized wear of the rollers, the stacking order after production, the weight limit of the overhead crane and the priorities of specific qualities. In the status quo, the planning of the duo rolling mill is done manually. Well-trained experts apply heuristics and decide for near-optimal schedules that combine several goals: minimizing the makespan, optimizing the temperature-curve of the furnace and optimizing the stacking order to improve the follow-up processes. As product variation increases steadily and additional targets (e.g., optimizing CO_2 emissions or total energy consumption) are introduced, the planning problem becomes more complex each year. Therefore, tools are needed to support the experts and to further optimize reaching the targets.

The function-based approach of the presented use-case consists of two main parts. In the first step, the uncertainty of the production process is handled. The plates of each order require specific thicknesses which are achieved by a varying number of rolling passes depending on e.g. material consistence or required quality. The estimation of the process time of an order at the duo rolling mill is thus a challenging task which needs to be tackled within the planning and scheduling approach. For overcoming this obstacle, the authors analysed different supervised ML approaches for predicting the process time at the duo rolling mill, of which the Gradient Tree Boosting approach showed the most promising results. Since the focus in this paper is on the planning approach in general and due to the limited pages, the prediction modelling and procedure is not explained in detail, the following tables and enumeration should give an overview of used input and basic outcome though. The features for prediction are obtained from order specifications and include information such as: quality, material, the number of plates, the required lengths, width or thicknesses, the required temperatures, the weights or the steel slab numbers.

After several pre-processing steps of cardinality analysis, outlier detection as well as feature selection, the following algorithms are compared: Regression Tree (RT), Linear Regression (LM), Random Forest Regression (RF), Support Vector Regression (SVR), Artificial Neural Network (ANN) and Gradient Boosting Machine (GBR). The error measures Mean Absolute error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Normalized Root Mean Squared Error (NRMSE) serve as evaluation metric for comparison. The results can be found in Table 1.

Table 1. Summary of error measures

Algorithm/ Accuracy	RT	LM	SVR	ANN	RF	GBR
MAE	16.4	14.3	18.0	14.4	13.6	12.7
MAPE	28.5	25.5	31.1	25.8	24.0	22.9
MSE	553	581	624	353	335	283
RMSE	23.5	24.1	25.0	18.8	18.3	16.8
NRMSE	13.4	13.8	14.3	10.7	10.6	9.6

The last column Gradient Tree Boosting shows best results with respect to all error measures and is thus currently the most promising approach for predicting the process time at the duo plant for planning and scheduling.

The second step of the function-based approach includes the optimization and the planning and scheduling of the orders. At first, the termination of orders is carried out on a weekly basis by the ERP System SAP. Under consideration of the planning horizon for detailed scheduling and sequencing, which in this use-case is defined by a setup event of the duo plant called roller change, the backlog for the next planning period can be defined. The scheduling optimization finally can be described by the following variables, restrictions, and objective functions.

Optimization variables for each order:

- Assigned shift,
- Sequence within a shift,
- Assigned stack.

Restrictions:

- No production after shift end,
- Sequence based on priorities of qualities,
- Sequence based on descending width classes of plates (avoiding rills and pits through outworn rollers),
- Sequence based on length of plates (building stable stacks after rolling).

Objectives:

- Energy consumption (approximated by the integral of the temperature curve),
- Number of offloading events of two available stacks.

The scheduler thus strives for identifying the optimal sequence for minimizing the energy consumption as well as the manual effort arising through the offloading and managing of stacks. These stacks need to be homogenously built and sorted for the subsequent heat-treatment process whereas each is limited to a maximum weight of 17 tons. Moreover, due to limitation in space the maximum number of stacks at a time equals two and for ensuring stability of the stacks, the dimensions of the plates need to be considered already in the scheduling approach. The dimensions of the plates also play an important role regarding the quality of the processed plates. In order to avoid pits and rills by outworn rolls orders are assigned to width classes based on the required width of the plates and the production sequence is built with respect to descending width classes. With other words, the optimizer is only allowed to alter the sequence of orders having the same width class. The energy consumption is approximated by the integral of the temperature profiles resulting from the required temperature levels of each order. The weighting between the two objective functions is accomplished by domain experts.

The optimization algorithm itself is based on an evolutionary algorithm consisting of the steps of selection, cross-over and mutation, as well as an elitism strategy and is implemented in Python. For selecting individuals from the parent generation, a probability distribution according to the fitness value is calculated. i.e. individuals with better fitness values are more likely to be selected. In the following cross-over step the genes of the individuals are combined resulting in a child individuum, with half of the gene information from each parent. Mutation finally randomly adapts single genes of an individuum such as, start time or coil setup of an order. Figure 4 shows a scheduling result obtained by the optimizer. In the first row each box corresponds to an order and its process duration on the duo mill. The grey boxes symbolize waiting time, which might be necessary to reach the desired temperature for a specific order. The colours of the boxes correspond to the temperature levels, whereas the temperature profile over time is depicted in the second row. The width of the plates are shown in the third row which are, apart from a few prioritized orders at the beginning, arranged in an descending order. The weight levels of the two stacks are finally depicted in rows 4 and 5.

Figure 4. Production schedule example

IV. CONSIDERATIONS ON THE PRACTICAL APPLICATION

In this chapter, we consider the integration of ML assisted planning approaches into real production environments. Unfortunately, most scientific articles focus only on the specific algorithms and the validation of the results and allude the next logical step: how to use the results and feed it back to the actual planning process. One reason for this is the shift of the scientific field from computer science/engineering to management. The other reason is that applications of the results is then performed as part of an integrated solution via software development and thus becomes out of scope for the articles at hand [14]. Fortunately, some high-level general hints can be extracted from the articles which can be summarised along two lines: *the workflow* and the *update rules*.

With regards to *the workflow*, there are two main approaches: 1) a human operator updates the planning parameters in the live system, 2) the integrated planning system updates the planning parameters autonomously. The tipping point between the two comes down to the number of parameters to be updated, the detail level of the models used during learning and the overall information system integration of the factory in question. Human operators tend to favour small number of parameters (~50) and beyond that, software support to update/accept many of them at once. On the model detail aspects, the need for manual adjustments stem from how the model was able to capture "quirks" or special corner cases of the underlying production. The more irregular the production process is the more likely the model needs to be overruled. Lastly, if the model ends up integrated with the information system the factory is run on, the application of the results is then done with no human operator involvement. Corner cases are either ignored, suppressed or reverted to more "classical" planning processes [15].

With regards to the *update rules*, there are three main approaches: 1) apply results verbatim, 2) alpha-blend old and new planning parameters, 3) human operator decides. Applying the results verbatim typically happens when the factory process isn't that complicated, and the algorithms produced high quality and reliable results. Usually, automatic application is preceded by a verification the parameters are within reasonable bounds. The drawback of this approach is that drastic (even if reasonably bound) changes can cause sudden drifts in the production process to the extent as if the factory was pulled on strings around. The alpha-blend method is more subtle in terms of introducing changes to the physical process by using some percentage of the previous parameters and some percentage of the newly learned parameters (i.e., 80%-20%) split. The advantage is the slow drift of the process towards the desired performance metrics. The disadvantage is the slower drift – the improvement may be spread across several dozen planning cycles. The last approach is the general human decision which considers the previous parameters, the learned parameters and external domain knowledge of the physical process not necessarily captured by the models. This approach may also involve the previous two approaches on a case-by-case basis. For example, the operator decides some parameters are applied verbatim, some with a 90%-10% blend, some with a specific value ignoring the learned parameters due to how the factory will have an out of ordinary period until the next planning cycle [16].

The last consideration is about the change in parameters and how the factory (shop floor) reacts to those changes. Most specifically, tightening operation times is typically met with resistance. There are valid reasons for not trying to perform certain operations too fast in general, no matter what the model learned suggests: equipment reliability, quality, safety, resilience and redundancy considerations in general. The other reasons are simply due human nature and willingness. These brings us to perhaps the largest caveat of applying machine-learned planning parameters in practice, the factory (shop floor) may try and act adversarial, counteracting the effects of the "forced upon" parameter changes. This can manifest, for example, in artificial process slowdowns so that the master data the ML is getting trained upon causes the resulting model to drift towards the original pre-ML planning parameters. Avoiding this caveat is more of a managerial task than an information system one; all involved parties of the manufacturing process (planners and workers) need to be heard and involved in the designing of the feedback approach and understand the shared goal and benefits to be achieved by this new planning process [17].

V. CONCLUSION AND OUTLOOK

The goal of the research project MLinPPC was to explore and implement different ML based planning approaches in order to systematically improve the PQ. In this publication we have summarized the results of this project. In the first part of the paper, we give an overview of the related literature and review previous work of the authors. We then discuss the implementation of two planning approaches, the evolutionary planning approach and the function-based planning approach, in a case study with two industry partners. While the results of the study demonstrate the applicability of both approaches, a comparison to suitable benchmarks, such as the currently used planning approaches, is still work in progress and thus a final evaluation is still needed. Moreover, in the first use case of the study, further development of the prediction models is needed as well. This includes the detection of special events that are not covered by the available training data or would potentially bias it. In the last part we considered the adoption of ML assisted planning approaches in production and the obstacles that exist to their adoption. While these obstacles usually get less focus in the literature, they may be crucial for success or failure of such methods in practice. Therefore, a well-defined strategy for deploying ML assisted planning approaches in production is essential. Such strategies should include a distinct workflow for production parameter and model updates, backup rules, in case of system delays, high errors or uncertainties, systems for error detection or uncertainty quantification and result communication to involved personal, such as shop-floor workers.

VI. ACKNOWLEDGMENTS

The authors would like to acknowledge the financial support of the EC for funding the H2020 research project EPIC under grant No. 739592. This research has been supported by the Austrian Research Promotion Agency (FFG) in the 32. "Produktion der Zukunft" Programm with the project title "MLinPPC - Increasing planning quality through the use of machine learning in production planning and control".

REFERENCES

[1] **Schuh, G., Schmidt, C.:** Produktionsmanagement, Springer Vieweg Berlin, Heidelberg, 2014.

[2] **Mather H., Plossl G.:** Priority Fixation Versus Throughput Planning, Journal of Production and Inventory Management, 19:27–51, 1978.

[3] **Gallina, V., Lingitz, L., Karner, M.**: A New Perspective of the Cyber-Physical Production Planning System *16th IMEKO TC10 Conference*, Berlin, Germany, 2019

[4] **Gallina, V., Lingitz, L., Breitschopf, J., Zudor, E., Sihn, W.**: Work in Progress Level Prediction with Long Short-Term Memory Recurrent Neural Network, *Procedia Manufacturing*, 54, pp. 136-141, 2021

[5] **Lingitz, L., Gallina, V., Ansari, F., Gyulai, D., Pfeiffer, A., Sihn, W., Monostori, L.**: Lead time prediction using machine learning algorithms: A case study by a semiconductor manufacturer, Procedia CIRP, 72, pp. 1051-1056,

[6] **Gyulai, D., Bergmann, J., Gallina, V., Gaal, A.**: Towards a connected factory: Shop-floor data analytics in cyber-physical environments, **Procedia CIRP**, 86, pp. 37-42, 2019

[7] **Gyulai, D., Pfeiffer, A., Bergmann, J., Gallina, V.**: Online lead time prediction supporting situation-aware production control, *Procedia CIRP*, 78, pp. 190-195, 2018

[8] **Lingitz, L., Gallina, V., Breitschopf,, J., Finamore, L., Sihn, W.**: Quality in production planning: Definition, quantification and a machine learning based improvement method, *Procedia Computer Science*, 217, pp. 358-365, 2023

[9] **Haug, Anders, Frederik Zachariassen, and Dennis Van Liempd.**: "The costs of poor data quality." Journal of Industrial Engineering and Management (JIEM) 4.2 (2011): 168-193.

[10] **von Rueden, L., Mayer, S., Sifa, R., Bauckhage, C., Garcke, J.:** Combining Machine Learning and Simulation to a Hybrid Modelling Approach: Current and Future Directions. In: Berthold, M., Feelders, A., Krempl, G. (eds) Advances in Intelligent Data Analysis XVIII. IDA 2020. Lecture Notes in Computer Science(), vol 12080. Springer, Cham, 2020

[11] **Ghobakhloo, Morteza, and Masood Fathi.** "Corporate survival in Industry 4.0 era: the enabling role of lean-digitized manufacturing." Journal of Manufacturing Technology Management, 2019.

[12] **May, Gökan, and Bojan Stahl**. "The significance of organizational change management for sustainable competitiveness in manufacturing: exploring the firm archetypes." International Journal of Production Research 55.15 (2017): 4450-4465.

[13] **Ryback, T., Lingitz, L., Gaal, A., Gallina, V., Gyulai, D.**: Improving the Planning Quality in Production Planning and Control with Machine Learning, *16th IMEKO TC10 Conference*, Berlin, Germany, 2019

[14] **Weichert, D., Link, P., Stoll, A., Rüping, S., Ihlenfeldt, S., & Wrobel, S.**: A review of machine learning for the optimization of production processes. The International Journal of Advanced Manufacturing Technology, 104(5-8), 1889-1902, 2019.

[15] **Messner, M., Pauker, F., Mauthner, G., Frühwirth, T., & Mangler, J.**: Closed loop cycle time feedback to optimize high-mix/low-volume production planning. Procedia CIRP, 81, 689-694, 2019.

[16] **Cimini, C., Pirola, F., Pinto, R., & Cavalieri, S.**: A human-in-the-loop manufacturing control architecture for the next generation of production systems. Journal of manufacturing systems, 54, 258-271, 2020.

[17] **Lauer, Thomas.** Change management. Springer Berlin Heidelberg, 2010.