2nd Conference on Production Systems and Logistics

Combining Process Mining And Simulation In Production Planning

Adrian Langer[1], Christian Ortmeier[1], Niels L. Martin[1,2], Tim Abraham[1], Christoph Herrmann[1]

[1]*Institute of Machine Tools and Production Technology / Chair of Sustainable Manufacturing and Life Cycle Engineering / Technische Universität Braunschweig, Langer Kamp 19b, 38106 Braunschweig, Germany*
[2]*Center of Pharmaceutical Engineering (PVZ), Technische Universität Braunschweig, Franz-Liszt-Str. 35A, 38106 Braunschweig, Germany*

Abstract

The conditions for industrial companies are changing due to increasing customer demands for individualised as well as sustainable products. Furthermore, companies are confronted with technological change by digital transformation. Therefore, production planning has to address various structural, procedural and organisational changes. Planning projects are often characterised by a high degree of complexity. In order to master the associated challenges, simulation models are used in production planning. In contrast to mathematical-analytical methods, simulation models examine and assess especially complex production systems and support improvement measures. A major difficulty during the model initialisation and the determination of the planning variables is the capture of data and the assurance of sufficient data quality. Both are associated with a high expenditure of time. At this point, manufacturing companies are faced with a conflict of objectives between the reduction of the planning time and the development of reliable simulation models. Process Mining (PM) can be used to capture data from central information systems and to uncover social and organisational networks and map them in a process model. This can create a well-founded data basis for simulation models.
To support simulation models within the planning process, a methodology linking process mining and simulation has been developed. This methodology improves the database within the planning process and renders it usable for rescheduling production systems. Potentials that can be achieved in the areas of data acquisition, data quality and model building are systematically analysed. The approach is validated on the basis of a use case from the pharmaceutical industry.

Keywords

Production Planning; Simulation; Process Mining; Data Acquisition

1. Introduction

Nowadays, production companies are confronted with decreasing batch sizes, highly fluctuating production rates, shorter product life cycles, shorter delivery times and an increasing number of product variants [1], [2]. To face these challenges, constant adjustments are necessary within the production system. This leads to an increasing effort in factory planning. For this reason, various approaches support the phases of factory planning by digital methods and tools [3], [4]. Simulations have proven to be a particularly valuable tool for decision support in factory planning. However, the data acquisition is associated with a very high effort. This results from the necessary data quality that must be available for accurate simulations. In this context, process mining (PM) offers the possibility to generate a data basis for simulations by analysing process-specific data.

For this purpose, past-oriented PM uses feedback data from existing IT systems. [5]
To support the initialisation of simulation models within the planning process, an approach that combines process mining and simulation is presented. The question is whether this approach can enable better simulation results and whether the effort in data acquisition and preparation can be significantly reduced. Therefore, the approach is applied in a case study of the pharmaceutical industry.

2. Theoretical background

2.1 Simulation models for decision support in factory planning

Factory planning encompasses the organisational, process-related and structural design or adaptation of the production system. Several approaches can be found in the literature that are based on a systematic, goal-oriented and phase-specific planning procedure that build on each other and combine the different planning tasks (e.g. process planning, layout planning, programme planning, etc.) [6], [7]. Basically, those planning phases develop from a draft to detailed planning and from an ideal to a real state. Starting from the definition of the goal, the further planning phases become more and more detailed. Subsequent planning phases can already begin even though previous ones have not yet been completed. In general, a distinction can be made between different basic planning cases [4], [6-8].
For each planning case, the classification in the planning level must be clarified. Starting from a rather global view for the network and site level, the following levels add more and more details reaching up to focus on the granular, detailed workplace of a specific area of a factory. Individual planning projects can either have an impact on several levels or be anchored in individual levels. The area structure level is responsible for the new planning or rescheduling of production chains and units within the production system. Here, the linking or arrangement of production stations is achieved via corresponding flow systems [3], [4], [6], [8]. The demands on the interdependencies and data also increase with increasing level of detail of the respective planning phases. With a high level of data availability and quality, more precise planning results can be achieved and the (re)planning project can be improved.

For further decision support, simulation tools can be used within the factory planning process. In general, models are an abstracted representation of the real system under investigation. In a manufacturing context, models can be classified based on the purpose of use. Jockisch and Rosendahl as well as Page and Kreutzer distinguish between descriptive, explanatory, forecasting and decision models. Besides the descriptive models, all other models are used in simulation modelling. Forecasting and decision models are the most common models in manufacturing for decision support. For simulation purposes, forecasting models use explanatory models to predict future scenarios. Decision models, which include optimisation models, have the purpose to facilitate the determination of options for action. For this purpose, the findings from an explanatory or forecasting model are supplemented by a decision space within the framework of decision models [9], [10].
In the context of factory planning, simulation has an important role with regard to the design and evaluation within the planning, realisation and operation of production systems to determine the operational performance. Depending on the planning case, the basic relationships of the production system already exist (brownfield) or are set up completely from scratch (greenfield). In rough planning, models can be initialised that have a high degree of abstraction, i.e. a low level of detail. As the planning process progresses, the demands on the models and thus also on the data and the associated level of detail grow. The biggest obstacle in creating models for real manufacturing applications is the difficulty in obtaining the right data and the determination of interrelationships [11]. The aim of modelling and simulation is to transfer conclusions gained through performance measurements on the model to the real system. In this way, simulation can support decision-making processes, which can assist in the selection and evaluation of solutions in the production system, especially in the case of several system variants related to the processual, organisational and structural interrelationships [12-14].

In addition, simulation models can be differentiated with regard to the paradigms chosen. The most important paradigms in simulation modelling are: System Dynamics (SD), Dynamic Systems (DS), Discrete Event (DE) and Agent Based (AB). Technically, SD and DS mostly deal with dynamic system behaviour, while DE and AB mostly deal with discrete system behaviour [15], [16]. In figure 1, typical simulation paradigms for specific elements of a production system are shown in relation to their degree of abstraction. In the context of factory planning, the necessary main elements have usually a higher degree of uncertainty and therefore a higher abstraction. According to this, the main elements for factory planning are highlighted in figure 1.

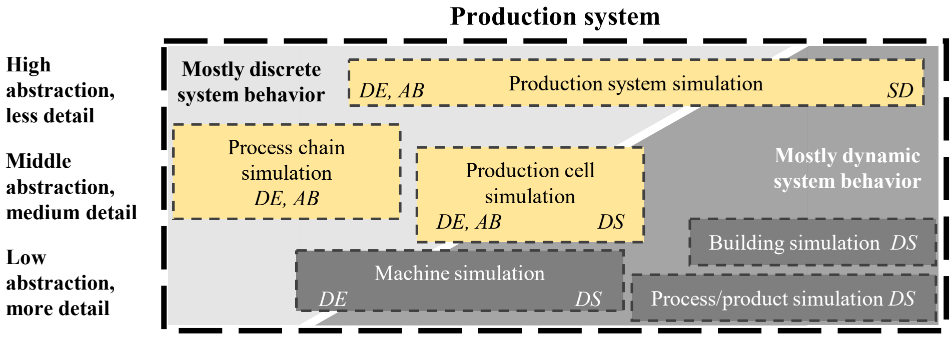

Figure 1: Overview of the different simulation models for different elements of a production system [16]

The definition of objectives in relation to the addressed planning tasks is particularly important in order to be able to select a suitable procedure for simulation. Material flows can be analysed via the discrete process chains by means of a DE or AB simulation, which changes the state only at certain events in a discrete manner. Ultimately, the approaches can also be combined depending on the planning task and objective. The combination of models for all elements of a production system, including product units, within a simulation would make it possible to analyse dependencies between the system elements involved and the effects of local improvement measures on the overall system [16]. In addition, a real-time simulation of the entire factory could run concurrently with real production operations and provide immediate results for short-term decisions [11]. However, a simulation is only as good as the input data. Process mining allows to access data of brownfield production systems.

2.2 Process mining

PM can be seen as a bridging technology between process science and data science. Process science includes, for example, techniques of workflow management, operations research and business process management. Within data science, besides techniques of statistics and data mining, also approaches of machine learning can be found. [17] PM techniques can be used to extract process-relevant information from an event log. These can generally be derived from Enterprise Resource Planning (ERP) or Manufacturing Execution Systems (MES). Event logs are a list of manual or system feedback data containing product and process specific information of a particular event. Van der Aalst et al. have defined five maturity levels for the data quality of an event log. Starting from the lowest maturity level, in which events only represent reality to a small extent and are increasingly paper-based, the highest maturity level records all events automatically and reliably. In addition to timeliness, the quality criterion of semantics is also met for all events [18].

By using algorithms, which are called miners in PM, a process model can be modelled based on the event logs. For this, an event log must contain at least a case ID, an event ID and a timestamp [18]. Furthermore, for a deeper evaluation of the process model, the description of the event (activity) as well as further order-, product- or process-specific information is useful. An example of event logs is shown in table 1.

Table 1: Event log of a data set

Case ID	**Event ID**	**Timestamp**	**Activity**	**Attributes**	**…**
Order number	Work station (WS)	Completion time	Process step	Quantity	…
010541528230	WS 01	14.12.2020 07:16	Preparation	7.000	…
010541528230	WS 02	15.12.2020 22:00	Production	7.000	…
010541528230	WS 03	15.12.2020 22:20	Post processing	7.000	…
010541528255	…	…	…	…	…

In process mining, van der Aalst distinguishes between the following types:

- *Process discovery* is used to create process models based on the event logs.
- *Process conformance* checks the degree of conformity between the process model and the actual observable process behaviour. The latter is described by the event logs.
- *Process enhancement* repeatedly links the created process model with event data to obtain further information, for example through performance analysis. In this way, errors in process models can be corrected or relevant decision points in the process can be focused [18].

Further, four different perspectives can be assumed in process mining, which can be the control-flow perspective, the organisational perspective, the case perspective and the time perspective [18]. Depending on the use case, different problems can be solved by using the types and perspectives of process mining. For the creation of a simulation model, for example, process discovery can be used to obtain a realistic model of the production processes. The purpose of this paper is to investigate the opportunities and obstacles of using process mining for simulations in the context of planning processes. For this, the current state of research is presented in the following.

3. Combination of PM and simulation

In literature, there are several methodologic approaches of combining process mining and simulation [5], [19], [20]. On the basis of the actual processes, Maruster and van Beest were able to determine the number of variants of the process flow as well as the average throughput times and incorporate these parameters into the simulation. The approach was applied in a Dutch government institution responsible for collecting fines and in the gas processing industry. In addition to this approach, the work of Aguirre et al. and Abohamad et al. also focuses on ensuring process transparency through PM. Thus, in both use cases, the actual process models could be described and averaged throughput times could be determined. The use cases were characterised by administrative tasks and activities. Aguirre et al. applied the developed process model to a procurement process in a university. Abohamad et al. used event logs in the healthcare sector for process analysis [5], [21]. The mentioned approaches are located in business process management and have delivered convincing results in this research field. Another approach by van der Aalst discusses the link between process mining and simulations and addresses the challenges faced by discovery techniques. Quality concepts such as recall, precision, and generalisation are focused here [22].
In contrast to these approaches, the focus of this work is on using PM for simulations to analyse an existing production system. The main difference is that target values from the work plan, such as processing and setup times, are already available for all process steps. In the manufacturing context, it can therefore be assumed that, due to a previous time recording, the planned processing times have a correspondingly high agreement with the actual processing times. By adding further influencing factors (e.g. machine failures, maintenance times, missing material), the results of the simulation should represent a well-founded image of reality. In the context of this work, it is therefore investigated whether improved simulation results can be achieved with data acquisition by PM.

4. Approach to combine process mining with simulations

The basic phases of a factory planning process are preparation, rough planning, detailed planning and implementation planning and execution. In the first phase, the goals and scope of the (re)planning project are derived. The actual simulation models can be used for rough and detailed planning to model the production system in order to gain knowledge about the production processes. Based on the results of the simulation, the final step is the implementation of the (re)planning project. The more accurately the results of a simulation match reality, the better an informed decision can be determined in the planning process. However, the more accurate the results, the more time and effort is required to collect and prepare the data. The simulation steps from rough to detailed planning are characterised for the planners by the comparison of effort and benefit. In practice, therefore, the effort is usually kept low and a simulation is generated on the basis of expert knowledge or planned values. Here, the approach aims at combining process mining with simulation, see figure 2.

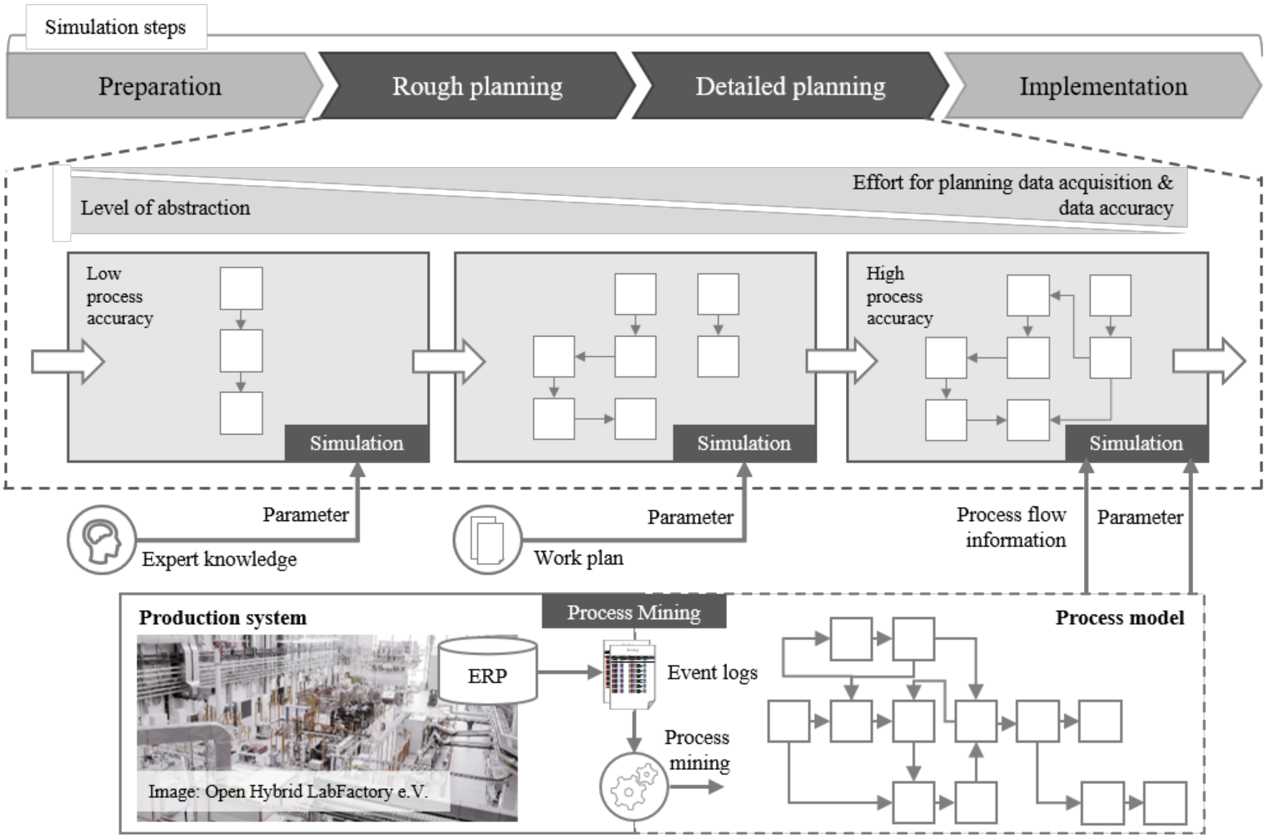

Figure 2: Approach to combine process mining with simulation

As shown in the approach, the PM can be used to create a process model based on the feedback data from the production system. In this context, the process model shows the actual production flow of manufacturing. This requires a correspondingly high level of data quality, and ensuring this is the core task of production data acquisition. Furthermore, parameters such as processing times, setup times, waiting times and failure times can be determined and used for the data basis. After an initial implementation of PM, these parameters can be continuously used for simulations with low effort and therefore, change the effort abstraction ratio for processes with high process accuracy. Furthermore, these parameters can be determined on a product-specific basis. In practice, product families are often formed in order to keep the effort low. Thus, in addition to reducing the effort, the data accuracy can also be increased. Hereby, it becomes clear that the support in the increasing data accuracy for the reduction of the associated efforts in the acquisition has a great impact on the planning processes. Ultimately, these conditions allow saving time as well as costs.

5. Use case

For this use case, real ERP order data for the final packaging of pharmaceutical products were received and will not be described in detail due to non-disclosure agreements. The data was cleared up of doublets as well as faulty datasets and finally comprises 912 orders for a whole production year (first half: 443; second half: 469) for four packaging lines. Besides the order specifics, the order data includes also the individually planned and real production time, such as process interruptions for each order produced on several process lines.

As illustrated in figure 3, the data is split into the first and second half of the year in order to allow the scenario of having the first half of the years' real production data and the planned data for the second half of the year. The process mining approach is applied on the production data and allows the reference with the second years' production data, which will be compared with the simulation data resulting from the input of the process mining data as well as the planned data. The comparison allows to conclude to what extent the planned data differs from the real production data compared to the use of PM.

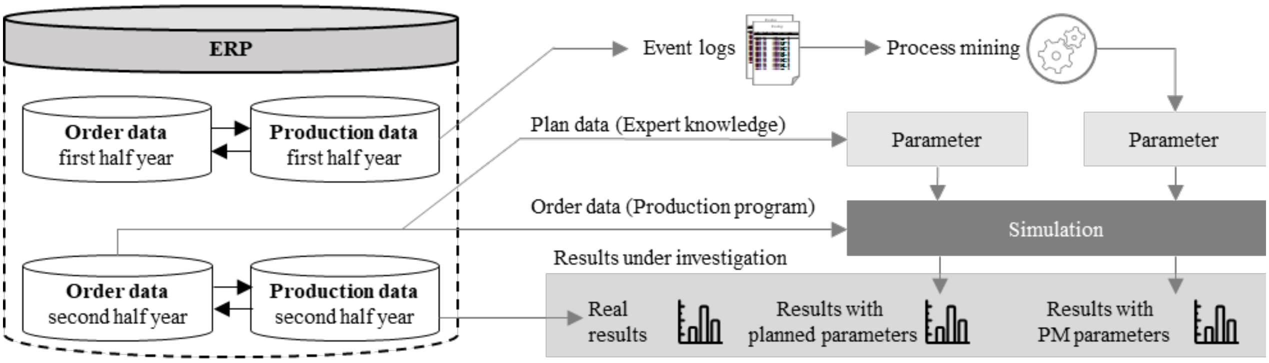

Figure 3: Comparison of real production data versus simulation scenarios by planned data and PM

A schematic overview of the agents and their interactions is illustrated in figure 4. The simulation is modelled in AnyLogic© using an agent-based simulation approach. The modelled agents of this model are a bulk storage for the raw, solid pharmaceutical products, the different process lines and an order scheduling. The bulk storage is able to dispatch products to the process lines and replenish the storage by defined amounts depending on the pharmaceutical product if an order from the process lines extends the stock of this product. The order scheduling assigns incoming orders to the process lines according to the lines' feasibility of handling these orders and the earliest possible production start at the order incoming date.

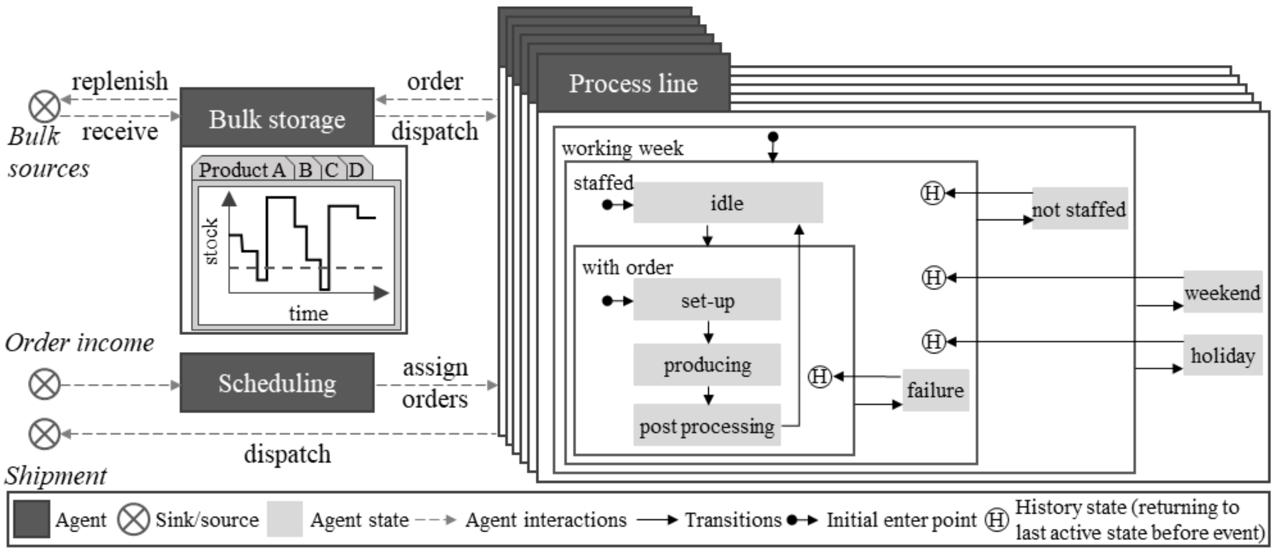

Figure 4: Schematic overview of the simulation model for the use case

The process lines consider weekends and public holidays and shift times to have a realistic and similar setting compared with the real production data on a macro scale. Those settings are the same in the three scenarios. On the micro scale, the process lines are in idle state until the order queue of the process line is filled and the order will be processed. If a process line processes an order, four states are possible: set-up, production, post processing and failure. For each state, there are parameters describing these states. Those parameters differ depending on the PM and the planning data and are listed in table 2.

Table 2: Applied process mining and planned data for simulation scenarios

State	Parameter	PM data [order data, 1. half]	Planned data [order data, 2. half]
Set-up	Set up time	Average set-up time of a specific process line	Planned set-up time of an order from order data
Production	Production ratio	The assessed production ratio of former orders per product group and line	The order's planned production rate
Post-processing	Post processing time	Average post processing time of a specific process line	Planned post processing time of an order
Failure	Failure times	Are included in production ratios	Are not considered in planning

The production ratio used in the PM and the planned data scenario are compared with the real order production ratio, which is shown in figure 5. The orders are sorted by rising planned production time. It can be observed that the real time data sets have outliers (A), which occur when the data logs are not logged properly for diverse reasons. Even though proper data sets are desirable, in practice, they are often hard to achieve and in this study they are not excluded. Furthermore, the production rates come closer to the real processes through data acquisition with the help of PM than can be achieved with the planning data. It can be observed that the real production rates for small and medium planned production rates are mainly too low, while the planned production rates for higher production rates match the real production rates better. This effect cannot be observed for the comparison of PM with the real production rates. Here, the real production rates seem to oscillate around the PM production rates with minor discrepancies.

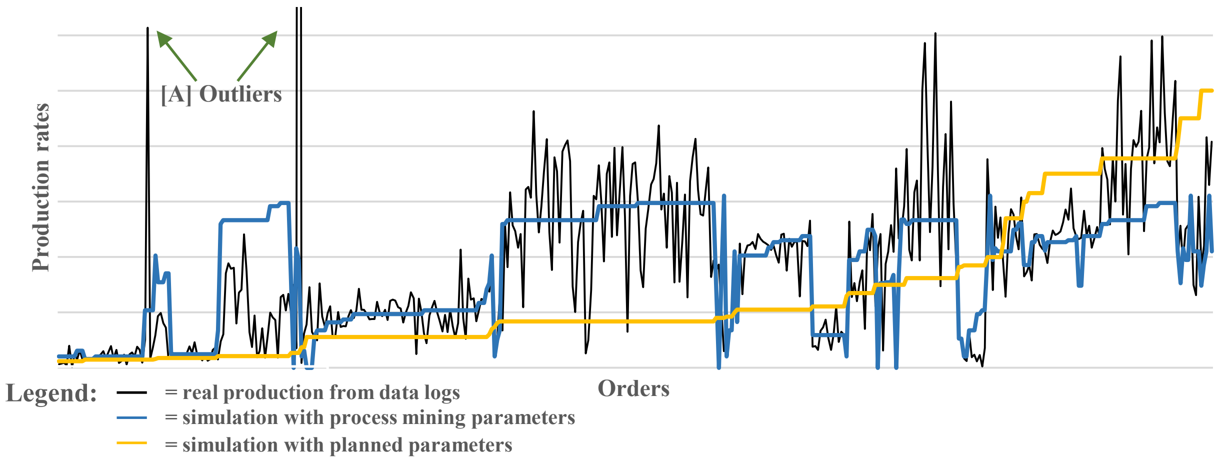

Figure 5: Comparison of PM and planned production ratios with the orders' real production ratio

This already shows that the process mining approach allows a better approximation of the input parameters for the simulation study. The diagrams in figure 6 show the deviations of the production volume and the throughput time from the simulation in relation to the real data. Regarding the above mentioned production ratio, the production volume shows a similar outcome. The real data sets and the data sets achieved by PM can be presented with a high amount of agreement. Here, however, the planned values are again found with

a higher deviation (figure 6, a). The middle area is particularly striking here; without offsetting the outliers, a very high agreement of the averaged production rates via PM compared to reality can be achieved. In comparison, the planned values are constantly below the real values. While for all three results the production volume is the same, it can be observed that the time needed for the completion of all orders is more accurate for the simulation results with PM in comparison to the planned data input. In this case, this leads to an overestimation of the required production time.
The diagram on the right side shows the cumulative deviation over time following the orders' chronological completion for the throughput time (figure 6, b). Here, it is particularly clear that the deviation increases as the number of orders increases for the simulation results using real and planned data. However, the deviation of the simulation results using real and PM data also exists, but the deviations compensate each other better. This effect is most probably a result of the usage of average production rates for product groups on specific production lines. Using a more specific approach might lead to even better results and should be a part of future research. Furthermore, these graphs allow the statement that the simulation itself is capable of forecasting the close to real throughput times using PM parameters.

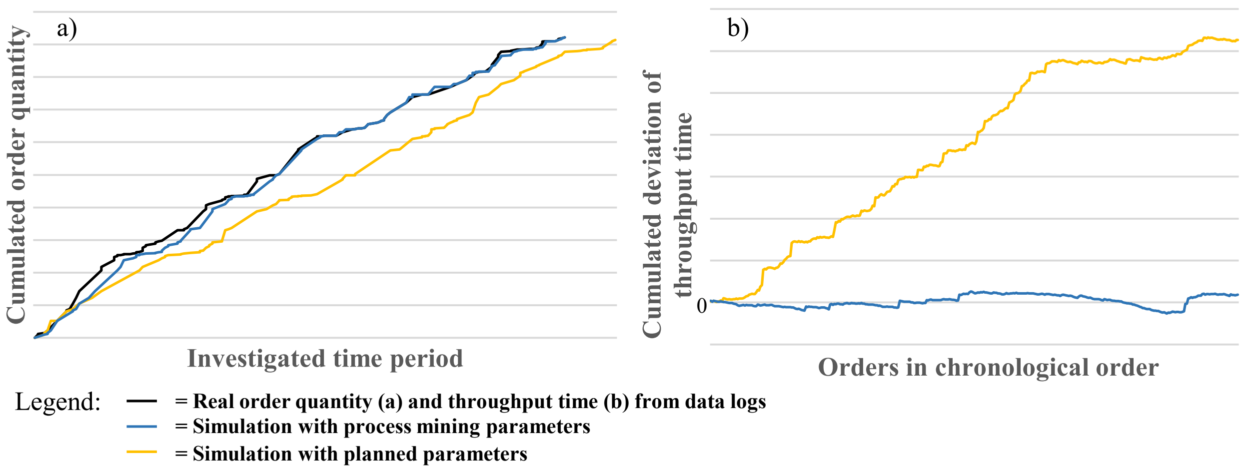

Figure 6: Comparison of real with PM and planned data for (a) the product volume and (b) the throughput time

Overall, this shows how strongly the planned values in this use case deviate from the real values. With PM, the parameters that serve as input for the simulation can be determined and transferred much more precisely. In the context of production planning, this is advantageous for the process-related, organisational and structural design.

6. Conclusion and outlook

In this work, the benefits of PM for ensuring a solid data basis for simulations were demonstrated. In an extensive use case of the pharmaceutical industry, process mining could be used for data acquisition and processing on the basis of an existing production line. The results demonstrated that, even in production planning, data is often not sufficiently accurate to develop a simulation and derive decisions based on it. In addition, the effort in data provision and procurement can be reduced by PM. Especially after the initial implementation of PM, the simulation can be adjusted continuously and with little effort. Due to the high prediction accuracy of processing and throughput times, new potentials also arise for the subject area of production planning and control. Here, for example, PM can be used in production control for targeted order release in order to increase adherence to schedules. In addition to rescheduling, data acquisition with PM can be used as a data basis for new facility planning projects within factory planning. Based on the process-related throughput times, production rates and downtimes, a more detailed target description including the scope of functions can be designed.

Acknowledgement

This paper evolved of the research project “Synus” (Methods and Tools for the synergetic conception and evaluation of Industry 4.0-solutions) which is funded by the European Regional Development Fund (EFRE | ZW 6-85012454) and managed by the Project Management Agency NBank.

References

[1] Gaal, A., Hofer, K., Ryback, T., Lingitz, L., Sihn, W., 2020. An agent-based approach to planning and controlling adaptable cell-oriented assembly systems. Procedia CIRP 93, 1158–1163.

[2] Tolio, T., Ceglarek, D., ElMaraghy, H.A., Fischer, A., Hu, S.J., Laperrière, L., Newman, S.T., Váncza, J., 2010. SPECIES—Co-evolution of products, processes and production systems. CIRP Annals 59 (2), 672–693.

[3] Pawellek, G., 2008. Ganzheitliche Fabrikplanung: Grundlagen, Vorgehensweise, EDV-Unterstützung. Springer-Verlag, Berlin, Heidelberg.

[4] Grundig, C.-G., 2018. Fabrikplanung: Planungssystematik - Methoden - Anwendungen, 6., neu bearbeitete Auflage ed. Hanser, München, 333 pp.

[5] Aguirre, S., Parra, C., Alvarado, J., 2013. Combination of Process Mining and Simulation Techniques for Business Process Redesign: A Methodological Approach, in: Cudré-Mauroux, P., Ceravolo, P., Gašević, D. (Eds.), Data-driven process discovery and analysis. International Symposium, SIMPDA 2012, Campione d’Italia, Italy, June 18-20, 2012, vol. 162. Springer, Berlin, pp. 24–43.

[6] Bauernhansl, T., 2020. Fabrikbetriebslehre 1: Management in der Produktion. Springer Vieweg, Berlin, 388 pp.

[7] Verein Deutscher Ingenieure, 2011. VDI 5200 Part 1: Factory planning - Planning procedures. Beuth Verlag GmbH.

[8] Schenk, M., Wirth, S., Müller, E., 2014. Fabrikplanung und Fabrikbetrieb: Methoden für die wandlungsfähige, vernetzte und ressourceneffiziente Fabrik, 2., vollständig überarbeitete und erweiterte Auflage ed. Springer-Vieweg, Berlin, Heidelberg, 832 pp.

[9] Jockisch, M., Rosendahl, J., 2010. Klassifikation von Modellen, in: Bandow, G., Holzmüller, H.H. (Eds.), „Das ist gar kein Modell!“. Unterschiedliche Modelle und Modellierungen in Betriebswirtschaftslehre und Ingenieurwissenschaften. Gabler Verlag / Springer Fachmedien Wiesbaden GmbH Wiesbaden, Wiesbaden, pp. 23–52.

[10] Page, B., Kreutzer, W., Gehlsen, B., 2005. The Java simulation handbook: Simulating discrete event systems with UML and Java. Shaker, Aachen, 502 pp.

[11] Fowler, J.W., Rose, O., 2004. Grand Challenges in Modeling and Simulation of Complex Manufacturing Systems. SIMULATION 80 (9), 469–476.

[12] Negahban, A., Smith, J.S., 2014. Simulation for manufacturing system design and operation: Literature review and analysis. Journal of Manufacturing Systems 33 (2), 241–261.

[13] März, L., Krug, W., Rose, O., Weigert, G., 2011. Simulation und Optimierung in Produktion und Logistik: Praxisorientierter Leitfaden mit Fallbeispielen. Springer, Heidelberg, 220 pp.

[14] Verein Deutscher Ingenieure, 2014. VDI 3633 Part 1: Simulation of systems in materials handling, logistics and production - Fundamentals. Beuth Verlag GmbH.

[15] Borshchev, A., Filippov, A., 2004. From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools. The 22nd International Conference of the System Dynamics Society.

[16] Schönemann, M., Bockholt, H., Thiede, S., Kwade, A., Herrmann, C., 2019. Multiscale simulation approach for production systems. Int J Adv Manuf Technol 102 (5-8), 1373–1390.

[17] van der Aalst, W., Damiani, E., 2015. Processes Meet Big Data: Connecting Data Science with Process Science. IEEE Trans. Serv. Comput. 8 (6), 810–819.

[18] van der Aalst, W., Adriansyah, A., Medeiros, A.K.A. de et al., 2012. Process Mining Manifesto, in: Daniel, F. (Ed.), Business Process Management Workshops. BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011; revised selected papers, vol. 99. Springer, Berlin, pp. 169–194.

[19] Măruşter, L., van Beest, N.R.T.P., 2009. Redesigning business processes: a methodology based on simulation and process mining techniques. Knowl Inf Syst 21 (3), 267–297.

[20] Rozinat, A., Mans, R.S., Song, M., van der Aalst, W., 2009. Discovering simulation models. Information Systems 34 (3), 305–327.

[21] Abohamad, W., Ramy, A., Arisha, A., 2017. A hybrid process-mining approach for simulation modeling, in: Chan, W.K., D'Ambrogio, A., Zacharewicz, G., Mustafee, N., Page, E.H., Wainer, G. (Eds.), WSC'17. Proceedings of the 2017 Winter Simulation Conference, December 3-6, 2017, Las Vegas, NV, vol. 117. IEEE; Omnipress, Piscataway, NJ, Madison, WI, pp. 1–12.

[22] van der Aalst, W., 2018. Process Mining and Simulation: A Match Made in Heaven! Proceedings of the 50th Computer Simulation Conference 50, 1–12.

Biography

Adrian Langer (*1993) studied industrial engineering with a specialisation in mechanical engineering at the TU Braunschweig. Since 2019, he has been working as a research assistant at TU Braunschweig in the sustainable production and factory systems working group of the Institute of Machine Tools and Production Technology (IWF) with a special focus on simulation of production systems.

Christian Ortmeier (*1992) studied industrial engineering with a specialisation in mechanical engineering at the TU Braunschweig. Since 2019, he has been working as a research assistant at TU Braunschweig in the sustainable production and factory systems working group of the Institute of Machine Tools and Production Technology (IWF) with a special focus on data-driven methods and tools in lean administration.

Niels Lasse Martin (*1988) studied industrial engineering with a specialisation in mechanical engineering at the TU Braunschweig. Since 2017, he has been working as a research assistant at TU Braunschweig in the sustainable production and factory systems working group of the Institute of Machine Tools and Production Technology (IWF) with a special focus on pharmaceutical production.

Dr. Tim Abraham (*1987) studied industrial engineering with a specialisation in mechanical engineering at the TU Braunschweig. Since 2014, he has been working at the IST, Fraunhofer Institute for Surface Engineering and Thin Films completing his PhD in 2019. Since 2020, he is head of the sustainable production and factory systems working group of the Institute of Machine Tools and Production Technology (IWF).

Prof. Dr.-Ing. Christoph Herrmann (*1970) is university professor for Sustainable Manufacturing & Life Cycle Engineering and co-director of the Institute of Machine Tools and Production Technology (IWF), Technische Universität Braunschweig as well as director of the Fraunhofer Institute for Surface Engineering and Thin Films IST since November 2018.