5th Conference on Production Systems and Logistics

Method For Semi-Automated Improvement Of Smart Factories Using Synthetic Data And Cause-Effect-Relationships

Vera Hummel[1,2], Jan Schuhmacher[1]

[1]*ESB Business School, Reutlingen University, Alteburgstr. 150, 72762 Reutlingen, Germany*

[2]*Department of Industrial Engineering, Stellenbosch University, 145 Banghoek Rd., 7600 Stellenbosch, South Africa*

Abstract

Smart factories, driven by the integration of automation and digital technologies, have revolutionized industrial production by enhancing efficiency, productivity, and flexibility. However, the optimization and continuous improvement of these complex systems present numerous challenges, especially when real-world data collection is time-consuming, expensive, or limited. In this paper, we propose a novel method for semi-automated improvement of smart factories using synthetic data and cause-effect-relations, while incorporating the aspect of self-organization. The method leverages the power of synthetic data generation techniques to create representative datasets that mimic the behaviour of real-world manufacturing systems. These synthetic datasets serve together with the cause-and-effect relationships as a valuable resource for factory optimization, as they enable extensive experimentation and analysis without the constraints of limited or costly real-world data. Furthermore, the method embraces the concept of self-organization within smart factories. By allowing the system to adapt and optimize itself based on feedback from the synthetic data, cause-effect-relationships, the factory can dynamically reconfigure and adjust its processes. To facilitate the improvement process, the method integrates the synthetic data with advanced analytics and machine learning algorithms as well as and the cause-and-effect relationships. This synergy between human expertise and technological advancements represents a compelling path towards a truly optimized smart factory of the future.

Keywords

Smart factories; synthetic data; cause-and-effect relationships; semi-automated improvement; self-organization; data-driven decision-making

1. Introduction

Globalized customer markets require companies to adapt to small production batch sizes and rapidly changing demands. Short innovation, technology, and product life cycles, as well as the need for cost-effective customized products, call for reconfigurable process paths [1], [2]. Smart, self-organized factories offer the potential to achieve flexible and adaptive production processes to solve arising problems in factory environments dealing with dynamic and rapidly changing boundary conditions and increasing complexity [3]. Self-organized factories are capable to perform a target-oriented (re-)structuring of the factory system to adapt to changing conditions without the need for a higher (hierarchical) decision-making level [4], [5]. To carry out these adaptations, autonomously acting system elements in the subsystems (like semi-autonomous working groups in assembly or autonomously mobile workstations) are necessary, which allow a targeted-oriented self-adjustment to enable an adaptation of the subsystem [6], [7]. The concepts of self-organization and autonomous control are closely linked to each other [5]. Self-organization deals with the

DOI: https://doi.org/10.15488/15310
ISSN: 2701-6277
publish-Ing.

autonomous emergence of ordered structures in dynamic, complex systems. It refers to the way in which order emerges from within a system and a system designs processual and systemic structures by itself [5], [8]. Autonomous control describes, in particular, processes of decentralized decision-making in heterarchical structures on the execution level [9], [10].

In a smart factory, interconnected manufacturing processes use data processing for early improvements. The continuous improvement process (CIP) includes problem analysis, solution generation, implementation, and checking, leading to new standards [11]. The philosophy behind CIP is proactive, using employee expertise to prevent problems and achieve significant improvements through continuous small changes over time. The analysis of product, process and factory data of including historical data combining with cause-effect relationships enables the identification of measures to achieve or even exceed defined production target dimensions by a semi-automatic improvement process using the self-organization capabilities.

2. Analysis of the state of science

To set the theoretical base the state of science in the fields of continuous improvement of smart factories, synthetic data generation, data analytics and cause-and-effect relationship analysis is examined.

2.1 Continuous improvement of smart factories

The continuous improvement process (CIP) involves small improvements in small steps and is one of the tools of lean management [12], [13]. In contrast to innovation, which is a top-down approach, CIP involves the employees [12]. In traditional CIP, digital tools assist processes, with improvement measures drawn from human expertise, promoting cross-functional collaboration among employees from various departments.

In the last decade a number of advanced digital technologies known as Industrie 4.0 were developed, which offer new approaches to managing complexity and improving productivity [14]. Integrating digital technologies in Industrie 4.0 with established lean management principles like CIP can optimize processes by enhancing transparency, networking, and automation, offering companies substantial potential. A study by BearingPoint [15] shows that the successful use of Lean Management 4.0 varies greatly depending on the industry. In a study of 50 manufacturing companies, 72% of decision-makers acknowledged the potential of lean management and digitalization for optimizing business processes. However, most companies are still in the early stages of implementing digital technologies within lean management methods [15]. There is also a trend towards developing tools that autonomously perform data analysis without human assistance [16]. Existing approaches (I) use especially human-based knowledge and experience for a manual or semi-automated transfer to new problems or (II) require large amounts of real data for an automated process using advanced analytics or machine learning to identify patterns [17], [18].

2.2 Synthetic data generation

Synthetic data is used when real data is challenging to access, lacks detail or scale, or is incomplete [19]. The three primary categories of synthetic data include dummy data, rule-based generated synthetic data, and synthetic data generated using artificial intelligence [20].

Dummy data, often made with a mock data generator, is randomly generated and lacks the characteristics, relationships, and statistical patterns of the original data, making it unrepresentative of the actual data [20]. Rule-based generated synthetic data, on the other hand, is produced by defining specific rules for data generation, which can be achieved using simulation tools [20], [21]. Within the realm of data farming, simulation methods are utilized to generate and analyse synthetic data to gain targeted insights [22], [23]. Rule-based synthetic data generation relies on precisely defining all features, relationships, and statistical patterns, with the quality of the generated data depending on the accuracy of the predefined rules, making it valuable when real data is unavailable [20]. Using artificial intelligence algorithms, synthetic data can be generated by training an AI model with real data to learn its features, relationships, and statistical patterns, and then the model creates entirely new data points that mimic the characteristics of the original dataset [20].

2.3 Data analytics

Data analytics is the scientific process of mathematically and logically transforming historical data into insights to explain the past and to achieve better future decision-making [24], [25]. With increasing maturity of analytical skills, descriptive (*What happened?*), diagnostic (*Why did it happen?*), predictive (*What will happen (with which probability)?*) and prescriptive analytics (*How can we make it happen?*) are distinguished, which all build on each other [24], [25], [26]. The starting point for analysis is set by the descriptive analytics by analysing what has happened based on statistics and instruments such as tables and charts organizing, summarizing and visualizing data [24], [27]. After descriptive analytics, diagnostic analytics is used to uncover patterns, root causes, and connections among attributes. It employs data-driven machine-learning algorithms like k-means clustering [24]. Diagnostic analytics analyses attribute correlations and interactions, such as process errors' influences and root causes of process time delays. In the predictive analytics step, methods and models are utilized to forecast errors and disruptions based on identified patterns within the dataset [24], [25]. In the final step of prescriptive analytics, simulation and optimization approaches are employed to positively impact the future by making targeted decisions, such as optimizing assembly processes to prevent errors like incorrectly mounted or forgotten components.

2.4 Cause-and-effect relationships (CER)

This work utilizes cause-and-effect relationships to identify interrelationships in a product lifecycle and improve production semi-automatically. These relationships indicate connections between variables, where changes in one variable cause changes in others, and genuine connections, rather than just close links, must be ensured during the identification process [24]. One process that builds on cause-and-effect relationships is quality problem-solving (QPS), which is used in the manufacturing industry. This process consists of problem definition, problem analysis, cause identification, solution generation and selection, and solution implementation and testing [28]. The data that exists in a QPS system includes the problem, the causes, different types of solutions and a solution evaluation. If the problems and causes are represented graphically, the result is a bipartite graph [29].

3. Research Method

The procedure of this work is based on Österle’s Design Science Research [30] with a focus on a practice-oriented research approach. The work begins with an analysis phase using mixed literature analysis, including snowballing, to establish theoretical foundations, analyse and evaluate literature aspects, and identify CER in a product lifecycle along with relevant data types. In the draft phase, the method is created based on the part solutions, generating synthetical data, CERs, data streams and fundamentals from literature as well as research already carried out. The verification was done against the requirements based on use case 1 for the suitability of the synthetically generated data. The validation to prove the achievement of the goal for application of the developed method was partly done by the use case 2 for a smart product and production.

4. Development of the method for semi-automated improvement of smart factories

The continuous improvement process (CIP) involves small improvements in small steps and is based on a bottom-up approach. The idea of developing this innovative method to enrich the CIP is based exactly on this principle, using operational data (bottom) and constantly processing it to make new improvements for the whole (top) factory system. The enrichment of the CIP (see Fig. 1) with AI in the PLAN segment of "problem/potential analysis" uncovers hidden relationships between variables, enabling early detection of problems or opportunities. If detailed data is unavailable, synthetic data generation methods may be used. CER are then established through diagnostic analytics to analyse the production system [27].

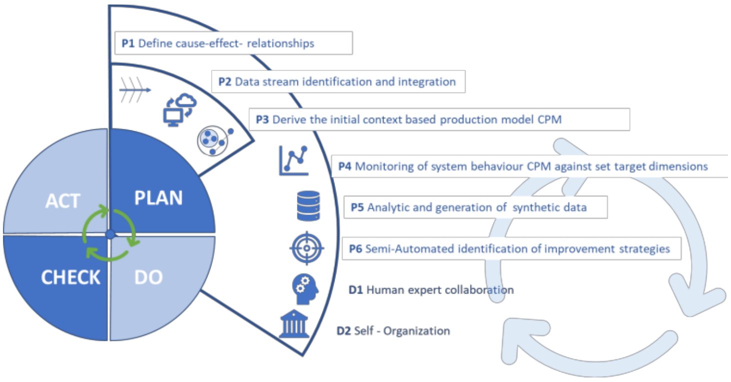

Figure 1: Enriched Continuous Improvement Process

The semi-automated measure generation through cause-and-effect relationships can contribute to increasing the efficiency of the CIP through digitalization. The partial solutions described below have been combined into a comprehensive method for semi-automated improvements. Smart factories, powered by data and AI, revolutionize industrial processes, boosting productivity and efficiency. To optimize them, a deeper understanding of cause-effect relationships is essential to spot potential improvements. Therefore, we propose a novel method that uses synthetic data and cause-effect relationships to create a production model for semi-automated improvements in smart factories (see Fig. 1).

The comprehensive method consists of: (P1) Utilizing Cause-Effect Relationships: Analysing critical factors in the smart factory environment; (P2) Implementation and Data Stream Integration: Testing improvement strategies in the real factory using continuous data streams; (P3) Development of a Context-Based Production Model: Building a comprehensive simulation model considering cause-effect-Relationships and the data streams; (P4) Monitoring of System Behaviour: Regularly observing and evaluating the production system against target dimensions for continuous improvement; (P5) Analytic and Generation of Synthetic Data: Using synthetic data to expand the data pool and conduct experiments safely; (P6) Semi-Automated Improvement Strategies: Employing AI algorithms to identify potential improvement strategies; (D1) Human Expert Collaboration: Factory personnel collaborate with AI to validate improvement strategies and (D2) Self-Organization: Enabling factories to autonomously adjust their structures and processes for optimal efficiency and adaptability.

4.1 Synthetic data generation

Synthetic data generation involves creating artificial data that resembles real-world data while excluding personally identifiable information or sensitive content. The main goal is to expand the dataset while retaining the statistical properties and patterns of the original data through methods like simulation studies. Following VDI 3633 seven steps have to be considered to perform a simulation study [31]. The procedure involves task definition, systems analysis, model formalization, implementation, simulation experiments, and result analysis. In this research the cause-effect-diagrams as well as the data streams will be used to formalize the model of the respective production system. Parallelly the phases of data collection and data preparation have to be considered. According to VDI 3633 [31], the mapping accuracy of the model should not be as detailed as possible, but as detailed as required to meet the specific objective.

For generating synthetical data a first minimal system model based structurally and process-wise on a real work system for the assembly of a product at Werk150, such as the MIMB box (see use cases 1 and 2), was designed and implemented [32], [33]. The major steps of the simulation study, including preparation, model implementation, and synthetic data generation as well as the results are described in chapter 5.1.

4.2 Cause-and-effect relationships

4.2.1 Development procedure of cause-and-effect relationships in general

The process for identifying and analysing problems and causes using Cause-and-Effect Diagrams (CED) and Detailed Cause-and-Effect Diagrams (DCED) are shortly described [29]. The first step involves identifying problems and causes through brainstorming and personal experience. Then, these problems and causes are clustered to prepare for the CED. The CED is created by categorizing main cause branches and identifying all possible causes. Causes are classified and analysed to understand their impact on the problem. The two-way relationship between problem class and cause class is determined based on predefined correlations and clustering. The DCED (Detailed Cause-Effect-Diagram) differs from the ACED (Abstract Cause-Effect-Diagram) as it only clusters problems [29]. The generation of a DCED is shown in Fig. 2. This process can be used to identify interrelationships in a product lifecycle, investigate potential problems and causes, and develop solutions. By understanding these interrelationships, processes can be derived to improve information flow and implement effective solutions, leading to problem elimination and prevention.

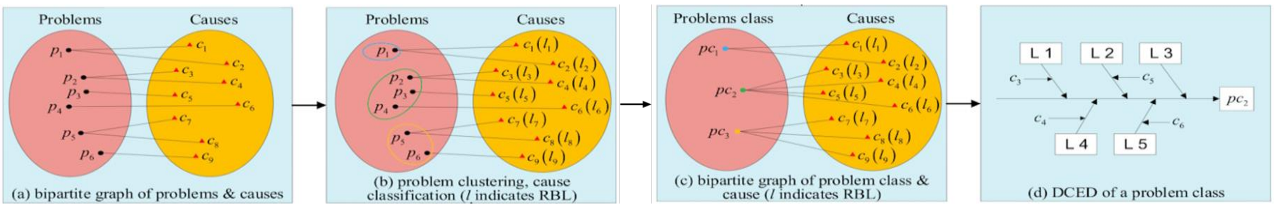

Figure 2: Generation of a DCED [29]

4.2.2 Definition of cause-and-effect relationships in process engineering and production

Based on literature [28], [30], [31], [32], [33], [34], [35], [36] and the above-described method Schnabel [37] defined cause-effect relationships for the process engineering as well as for production, structured them into classes, documented them into detailed cause-and-effect diagrams (see Fig. 3) for defined problem and challenge classes [38]. Both will be used for the design of the context production model CPM.

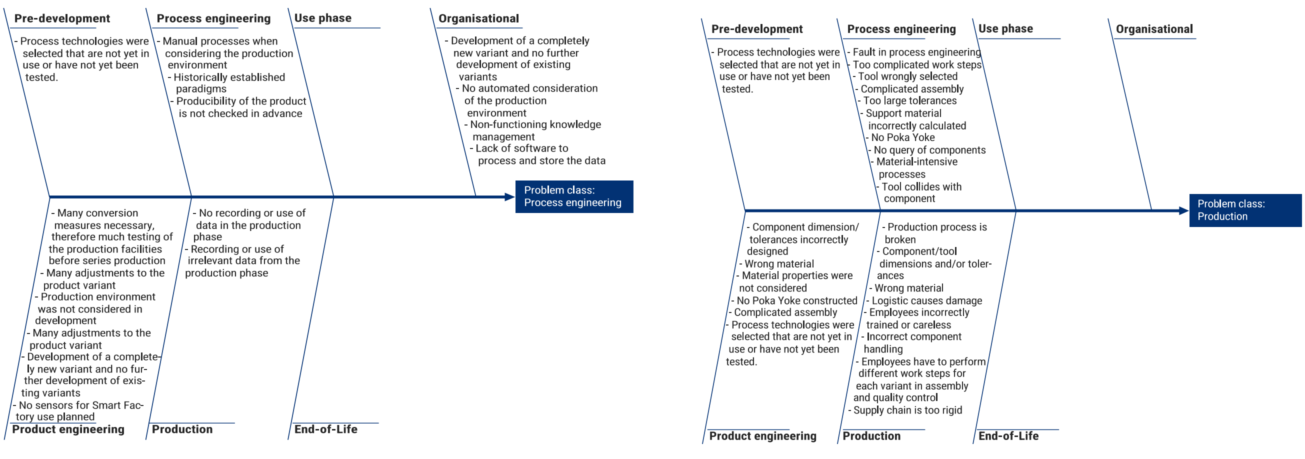

Figure 3: Cause-and-effect-diagram for problem class "process engineering" and “production” [37]

4.3 Data streams in process engineering and production

To identify relevant data in the product lifecycle, particularly for process engineering and production, the procedure developed by Günther et al. [39] is employed. Process analyses are recommended to identify relevant influencing factors and problem sources. Cause-and-effect relationships are used to create an overview of data streams in the product lifecycle. Relevant data sources are selected and evaluated based on criteria, including relevance, reliability, availability, data stream characteristics, and customer data security. In the next step additionally required data sources are determined based on systems’ requirements [37]. The relevant data type was assigned to the product life cycle phase and as input data required or as output generated. The integrated product engineering model (iPeM) [38] with the main phases pre-development, product engineering, process engineering, production, usage and end of life builds the basis. During process engineering, product data from previous phases and subsequent generations is utilized. This includes information on product properties, production capability factors, production processes, quality data, use phase data, and remanufacturing processes. In the production phase, data from product and process engineering is essential. To achieve continuous improvement, processes must be implemented based on data collected during series production, following the concept of self-organization. The data types developed are described and documented in Schnabel [37].

5. Validation

Validation is the process of checking whether the proposed solution meets the expected objectives. Verification (UC1) is to check if the requirements for the usage of the synthetically generated data are fulfilled. Here, the correctness of the overall result is validated [40]. To reuse the partial solution for future challenges, validation tests with use cases (see Fig. 4, UC2 and UC3) that interconnect on each other were built.

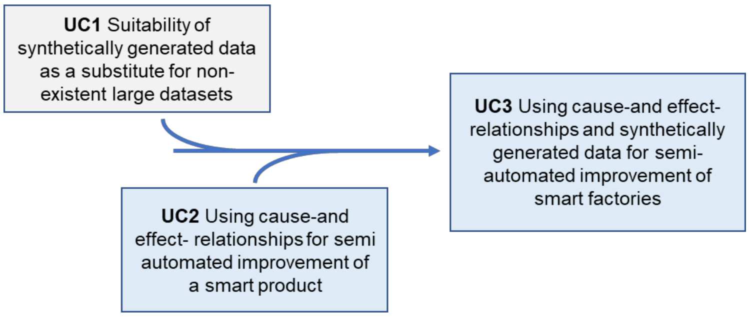

Figure 4: Verification and Validation approach

The chosen method for validation is experimentation where a specific assumption or conjecture is validated in a practice-oriented investigation.

5.1 Use case 1: Synthetic data generation

Use case 1 dealt with the investigation of the suitability of synthetically generated data of production and process data of a defined minimal system (see Fig. 5) with reduced structural complexity for a data-driven prediction of turbulence events and their effects on process and throughput times [33].

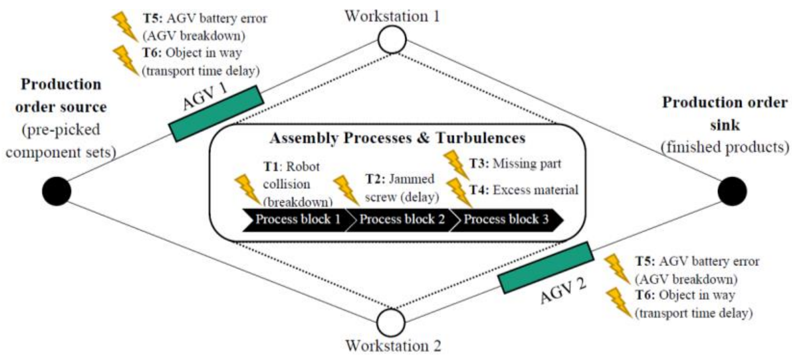

Figure 5: Defined minimal system with turbulence events (T) [33]

The minimal system model is based structurally and process-wise on an actual work system for the assembly of a product at two alternative workstations which are supplied with pre-picked component sets from the warehouse via automated guided vehicles (AGV). For the generation of the synthetic data set a rule-based simulative approach has been applied, since no real data on production systems' internal turbulences was available, but the effects of the turbulences on the production system was known. The simulation study to generate the synthetic data has been done following the procedure of VDI 3633 [31]. For the modelling of the defined minimal system, a multi-method simulation with agent-based modelling of the transport resources' autonomous behaviour and event-oriented modelling of the turbulence scenarios with stochastic turbulence events has been applied using the simulation tool AnyLogic. Detailed information about the considered turbulences, turbulence attributes, and considered data, values and distributions as well as about the entire simulation study can be found in Schuhmacher [33]. For the investigation of suitable data-driven prediction methods the generated synthetic data has been divided into training data (60 %) for training the data-driven prediction models and test data (40 %) for determining the prediction accuracies. The considered models and methods of different types (Naïve Bayes, generalized linear model, logistic regression, fast large margin, deep learning, decision tree, random forest, gradient boosted trees) have been compared in terms of prediction accuracy for the respective turbulence, standard deviation and computation time. Based on this use case it has been shown and validated that a data-driven, probabilistic prediction can be achieved through a simulative generation of synthetic data. To transfer the identified potentials of the synthetic data generation to higher-level questions of the CIP despite the lack of real data, the developed "Enriched Continuous Improvement Process" (see Fig. 1) will be applied. By building up a simulation model (phase P3), monitoring the system behaviour (phase P4), generating and analysing the synthetic data (phase P5) a semi-automated identification of improvement strategies (phase P6) to reach the set targets should be achieved, incorporating expert knowledge of the employees (phase D1).

5.2 Use case 2: Using CER for semi-automated improvement of a smart product

The use case 2 was carried out in the smart factory Werk150, the factory of the ESB Business School on the Campus of the Reutlingen university as part of a comprehensive research project [37]. The Werk150 is a research and education environment with innovative industrial digital engineering and physical manufacturing infrastructure, a so called "Industry 4.0 Factory". Werk150 is a research environment where the latest production and logistics technologies are developed and represent a state-of-the-art CPS production system. To develop and test technologies a city scooter and a modular, intelligent mobility box is manufactured [32]. To define the limitations of the use case 2 the system to be investigated is described according to [41]. In this validation, the system to be examined is a product, which is developed with the i2PLM see [37]. The input variables that can be changed are in the engineering phase the construction of the product depending on use phase data. In the production phase, the 3D printing parameters can be adjusted based on the product quality data. The input variables which are non-changeable are the material of the product and the production process, with its work steps.

Figure 6: Smart lunchbox

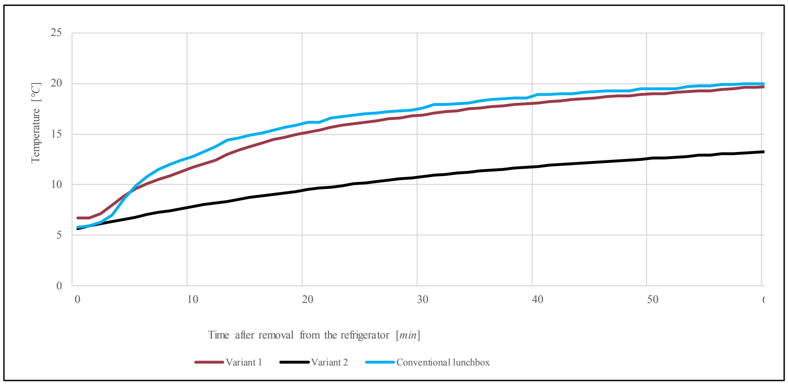

Figure 7: Temperature curve comparison of smart lunchbox Variant 1 and 2

The modular intelligent mobility box (MIMB) by Werk150 is a smart product (see Fig.6), with sensors for data storage and analysis. It connects to a network for data transmission and includes a 3D-printed top and bottom, a temperature and humidity sensor, data gateway, and velcro. The lunchbox's use case is to keep chilled food fresh during the commute, maintaining the temperature below 15 °C for one hour after removal from the refrigerator, with the sensor monitoring the internal temperature. The procedure for the experiment is based on the processes of the i²PLM and is shown in Fig. 8. In Variant 1 the lunchbox is already developed, and the process engineering has already been carried out. This is followed by the production of the lunchbox in Werk150 where data is generated on the quality characteristic. The product is used by experiment participants in the use phase where temperature data is generated while carrying out the use cases. These data are converted into a temperature curve which flows into a cause- and effect-relationship (CER) and triggers product and production changes for the subsequent generation (Variant 2). Variant 2 is a customized lunchbox. The changes in the product engineering of Variant 2 are mainly incorporated by changing the wall thickness, as this significantly influences the temperature curve on the product side. In process engineering, new requirements result in changes of 3D printing properties. Subsequently, the product is engineered, produced, and used and the same data types are recorded to validate product improvements.

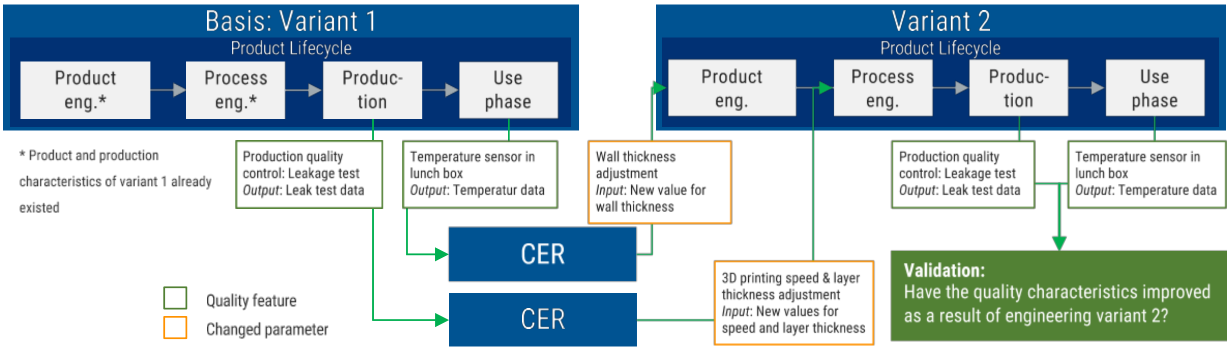

Figure 8: Schematic diagram of the experiment procedure

The results from production quality control are that Variant 2 is in contrast to Variant 1 leak-proof, and in addition, the surface is smoother, and the rework effort of the 3D printing is reduced. In the use phase, temperature curves are generated by 36 users and shown in the diagram in Fig. 7. The average temperature curve of Variant 1 is shown with a dotted line and Variant 2 with a solid line. Variant 2 meets the required temperature after one hour with 13.2 °C and improved by 6.5 °C compared to Variant 1. Improvements occurred in the product properties, as well as in the production time and quality. Use case 2 proofed that using the CER combined with the data streams the semi-automatic provision of the improvement potentials work in this use case quite well and will be the basis for the use case 3 for the validation of the comprehensive methods as presented in this paper.

6. Conclusion and outlook

The developed novel method for semi-automated improvement of smart factories using synthetic data and cause-effect relationships presents a promising approach to enhancing production efficiency and productivity. Both use cases have separately proven that the objective can be achieved. By leveraging the power of artificial intelligence, synthetic data, and cause-effect insights, factories can achieve intelligent decision-making and drive continuous improvements in their operations. The next steps for use case 3 (see Fig. 4) will involve the combined and integrated use of cause-and-effect-relationships and synthetically generated data for the semi-automated improvement of smart factories on the example of the complete production system of Werk150 to prove the path towards a truly optimized factory of the future. This will lead to deeper insights on the potentials and limitations of using synthetic data and cause-effect relationships for the semi-automated improvement of smart factories.

References

[1] Shi, Z., Xie, Y., Xue, W., Chen, Y., Fu, L., Xu, X., 2020. Smart factory in Industry 4.0. Syst Res Behav Sci 37 (4), 607–617.

[2] Arnold, C., Kiel, D., Voigt, K.-I., 2017. Innovative Business Models for the Industrial Internet of Things. Berg Huettenmaenn Monatsh 162 (9), 371–381.

[3] Hozdić, E., 2015. Smart factory for industry 4.0: A review. International Journal of Modern Manufacturing Technologies 7 (1), 28–35.

[4] Stock, D., Bauernhansl, T., Weyrich, M., Feurer, M., Wutzke, R., 2020. System Architectures for Cyber-Physical Production Systems enabling Self-X and Autonomy, in: 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA). 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria. 08.09.2020 - 11.09.2020. IEEE, pp. 148–155.

[5] Schuhmacher, J., Hummel, V., 2020. Development of a catalogue of criteria for the evaluation of the self-organization of flexible intralogistics systems. Procedia CIRP 93, 832–837.

[6] Wehberg, G.G. von, 2016. Logistik 4.0 – die sechs Säulen der Logistik in der Zukunft, in: Göpfert, I. (Ed.), Logistik der Zukunft - Logistics for the Future. Springer Fachmedien Wiesbaden, Wiesbaden, pp. 319–344.

[7] Holland, J.H., 2002. Complex Adaptive Systems and Spontaneous Emergence, in: Curzio, A.Q., Fortis, M. (Eds.), Complexity and Industrial Clusters. Dynamics and Models in Theory and Practice. Physica-Verlag HD, Heidelberg, pp. 25–34.

[8] Vec, M. (Ed.), 2006. Selbstorganisation: Ein Denksystem für Natur und Gesellschaft. Böhlau, Köln, 447 pp.

[9] Philipp, T., de Beer, C., Windt, K., Scholz-Reiter, B., 2007. Evaluation of Autonomous Logistic Processes – Analysis of the Influence of Structural Complexity, in: Windt, K., Hülsmann, M. (Eds.), Understanding autonomous cooperation and control in logistics. The impact of autonomy on management, information, communication and material flow. Springer, Berlin, New York, pp. 302–324.

[10] Windt, K., Hülsmann, M. (Eds.), 2007. Understanding autonomous cooperation and control in logistics: The impact of autonomy on management, information, communication and material flow. Springer, Berlin, New York, 417 pp.

[11] Klanitz, T., 2023. Kontinuierlicher Verbesserungsprozess (KVP). REFA AG. https://refa.de/service/refa-lexikon/kontinuierlicher-verbesserungsprozess-kvp. Accessed 24 October 2023.

[12] Helmold, M., 2021. Kaizen, Lean Management und Digitalisierung. Springer Fachmedien Wiesbaden, Wiesbaden.

[13] Ohno, T., 1988. Toyota production system: Beyond large-scale production. Productivity Press, Cambridge, Mass., 143 pp.

[14] Bauernhansl, T., 2017. Die Vierte Industrielle Revolution – Der Weg in ein wertschaffendes Produktionsparadigma, in: Vogel-Heuser, B., Bauernhansl, T., Hompel, M. ten (Eds.), Handbuch Industrie 4.0 Bd.4. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–31.

[15] Kettler, D., König, C., 2017. Lean 4.0 – Schlank durch Digitalisierung. BearingPoint. https://www.bearingpoint.com/files/Studie_Lean_4.0_Schlank_durch_Digitalisierung.pdf. Accessed 2 October 2023.

[16] Oztemel, E., Gursev, S., 2020. Literature review of Industry 4.0 and related technologies. J Intell Manuf 31 (1), 127–182.

[17] Yska, S., Bustos, D., Guedes, J.C., 2023. Machine Learning Applications for Continuous Improvement in Integrated Management Systems: A Short Review, in: Arezes, P.M., Baptista, J.S., Melo, R.B., Castelo Branco, J., Carneiro, P., Colim, A., Costa, N., Costa, S., Duarte, J., Guedes, J.C., Perestrelo, G. (Eds.), Occupational and Environmental Safety and Health IV, vol. 449. Springer International Publishing, Cham, pp. 541–551.

[18] Mjimer, I., Aoula, E.-S., Achouyab, E.H., 2023. Contribution of machine learning in continuous improvement processes. JQME 29 (2), 553–567.

[19] Mannino, M., Abouzied, A., 2019. Is this Real?, in: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology. UIST ‘19: The 32nd Annual ACM Symposium on User Interface Software and Technology, New Orleans LA USA. 20 10 2019 23 10 2019. ACM, New York, NY, USA, pp. 549–561.

[20] Syntho, 2022. Was sind synthetische Daten? Syntho. https://www.syntho.ai/de/what-is-synthetic-data/. Accessed 24 October 2023.

[21] AnyLogic, 2023. Synthetic data generation. AnyLogic. https://www.anylogic.de/features/artificial-intelligence/synthetic-data/. Accessed 21 October 2023.

[22] Lechler, T., Sjarov, M., Franke, J., 2021. Data Farming in Production Systems - A Review on Potentials, Challenges and Exemplary Applications. Procedia CIRP 96, 230–235.

[23] Sanchez, S.M., 2020. Data Farming. ACM Transactions on Modeling and Computer Simulation 30 (4), 1–30.

[24] Hill, R., Berry, S., 2021. An Introduction to Industrial Analytics, in: Hill, R., Berry, S. (Eds.), Guide to Industrial Analytics. Springer International Publishing, Cham, pp. 3–20.

[25] Schuh, G., Blum, M., Reschke, J., Birkmeier, M., 2016. Der Digitale Schatten in der Auftragsabwicklung. ZWF 111 (1-2), 48–51.

[26] Gartner, Inc., 2014. Gartner Says Advanced Analytics Is a Top Business Priority. Gartner, Inc. https://www.gartner.com/en/newsroom/press-releases/2014-10-21-gartner-says-advanced-analytics-is-a-top-business-priority. Accessed 24 October 2023.

[27] Olson, D.L., Lauhoff, G., 2019. Descriptive Data Mining, 2nd ed. 2019 ed. Springer, Singapore, 130 pp.

[28] MacDuffie, J.P., 1997. The Road to “Root Cause”: Shop-Floor Problem-Solving at Three Auto Assembly Plants. Management Science 43 (4), 479–502.

[29] Xu, Z., Dang, Y., 2020. Automated digital cause-and-effect diagrams to assist causal analysis in problem-solving: a data-driven approach. International Journal of Production Research 58 (17), 5359–5379.

[30] Österle, H., Becker, J., Frank, U., Hess, T., Karagiannis, D., Krcmar, H., Loos, P., Mertens, P., Oberweis, A., Sinz, E.J., 2010. Memorandum zur gestaltungsorientierten Wirtschaftsinformatik. Schmalenbachs Z betriebswirtsch Forsch 62 (6), 664–672.

[31] Verein Deutscher Ingenieure, 2014. Simulation of systems in materials handling, logistics and production: Fundamentals. Beuth Verlag GmbH, Berlin ICS 03.100.10.

[32] Reutlingen University - ESB Business School, 2023. Werk150: Learning, research and development environment - ESB Business School’s factory. Reutlingen University - ESB Business School. https://www.esb-business-school.de/en/research/research-activities/werk150. Accessed 24 October 2023.

[33] Schuhmacher, J., Bauernhansl, T., 2023. Data-driven Prediction of Internal Turbulences in Production Using Synthetic Data. Hannover : publish-Ing.

[34] Beckschulte, S., Kiesel, R., Schmitt, R.H., 2021. Manuelle Fehleraufnahme bei Mass Customization. ZWF 116 (4), 188–192.

[35] Klanitz, T. Poka Yoke. REFA AG. https://refa.de/service/refa-lexikon/poka-yoke. Accessed 2 October 2023.

[36] Rauch, E., Rojas, R., Dallasega, P., Matt, D.T., 2018. Smart Shopfloor Management. ZWF 113 (1-2), 17–21.

[37] Schnabel, B., 2023. Development of a Comprehensive Product Lifecycle based on Data of Smart Factories and Smart Products. Master Thesis.

[38] Gernhardt, B., Vogel, T., Hemmje, M., 2018. Knowledge-Based Production Planning for Industry 4.0, in: Hoppe, T., Humm, B., Reibold, A. (Eds.), Semantic Applications. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 181–202.

[39] Günther, R., Beckschulte, S., Wende, M., Schmitt, R.H., 2022. Wertschöpfungskettenübergreifendes Fehlermanagement. ZWF 117 (4), 187–191.

[40] Fenz, S., Ekelhart, A., 2011. Verification, Validation, and Evaluation in Information Security Risk Management. IEEE Secur. Privacy Mag. 9 (2), 58–65.

[41] Siebertz, K., van Bebber, D., Hochkirchen, T., 2017. Statistische Versuchsplanung: Design of Experiments (DoE), 2. Auflage ed. Springer Vieweg, Berlin, Heidelberg, 508 pp.

Biography

Vera Hummel (*1962) has been a professor at the ESB Business School (Reutlingen University) since 2010. She is Vice-Dean Research of the faculty and the initiator and head of the "Werk150 – The Factory of ESB Business School" for research, education and industry training on the campus of Reutlingen University. She is elected president of the International Association on Learning Factories IALF and holds an extraordinary professorship at the Stellenbosch University, Industrial Engineering department.”.

Jan Schuhmacher (*1987) has been a research associate at the ESB Business School (Reutlingen University) since 2014, involved in the establishment of the Industrie 4.0 Factory - Werk150 since 2015 and doctoral student at the Institute of Industrial Manufacturing and Management IFF at the University of Stuttgart since 2020. His major field of research is the design and optimization of changeable, autonomously controlled intralogistics systems.