Simulation in Produktion
und Logistik 2021
Jörg Franke & Peter Schuderer (Hrsg.)
Cuvillier Verlag, Göttingen 2021

Continuous Validation and Updating for High Accuracy of Digital Twins of Production Systems

Kontinuierliche Validierung und Aktualisierung für hohe Realitätsnähe von Digitalen Zwillingen von Produktionssystemen

Leonard Overbeck, Arthur Le Louarn, Oliver Brützel, Nicole Stricker, Gisela Lanza,
wbk Institute of Production Science, KIT, Karlsruhe (Germany),
leonard.overbeck@kit.edu, arthur.lelouarn@student.kit.edu,
oliver.bruetzel@kit.edu, nicole.stricker@kit.edu, gisela.lanza@kit.edu

Abstract: Despite continuous improvements in modelling, software tools and data availability, simulation projects of production systems still require a lot of manual effort, expertise in various disciplines and time. In many projects the high initial invest for building the simulation model is followed by a rather short period of experimentation and analysis. As production systems have to be adapted at an increasing pace to respond to rapidly changing markets and business environments, simulation models of these systems become outdated earlier, reducing their useful time window. One way to extend this time window would be the implementation of a method of automated comparison with the current production systems and subsequent self-adaption of the model to reality to maintain and even improve its accuracy over time. This approach will be presented and validated at a real world use case. Such an enhanced simulation model can be called a digital twin of the production system.

1 Introduction

Discrete-event simulation models (DES) permit the in-depth analysis and evaluation of improvement ideas on existing production systems without having to interfere with running production, which makes them a powerful tool for efficiency improvement of production (Mayer et al. 2020). Yet, in most companies simulation models of production systems are still built and used only in temporary projects (VDI, 2014). This leads to limited benefits by high initial costs, since simulation models require a lot of expertise and time to be creates and implements and even more to obtain satisfying accuracy. A longer usability would improve the return on investment of simulation models. But once a model is created, it constantly has to be adapted to changes in the real production system, if it shall be used over the whole life cycle of the production system for ongoing analysis and improvement. Since manual adaption is extremely time consuming, an approach of continuous validation of simulation models and automated updating was developed. Validation is by VDI (VDI, 2014) defined as the "examination of the model as to whether the real behaviour of the modelled system is sufficiently well rendered with regard to the examination target" (part 1, p. 21). The continuous validation and update from real production data turn the simulation model into a real digital twin of the production system (Kritzinger et al., 2018).

2 Literature review

2.1 Modell generation and maintenance

Splanemann (1995) was one of the first to try semi-automated simulation model generation. His approach primarily uses CAD data in STEP-format (STandard for the Exchange of Product model data) to model the layout of the production system automatically. Focussing more on model parameters, Werner and Weigert (2002) proposed an approach to parametrize a model template, which was developed a-priori by experts, with data from ERP (Enterprise-Resource-Planning) and PDA (Production Data acquisition) systems and performed an analysis of model convergence to reality.

Bergmann and Straßburger (2020) presented different tools and methods to automatically generate simulation models which help designing a high-automated update process. One important step in this research field is the dissertation of Bergmann (2013) which uses the Core Manufacturing Simulation Data standard to create simulation models. Kotiades (2016) introduced the concept of a Self-Adaptive Discrete Event Simulation (SADES) but did not provide an exemplary implementation.

A recent and more elaborate overview of existing approaches is given by Reinhardt et al. (2019).

2.2 Data input for simulation models

Robinson and Perera (2002) provide an early discussion of chances and obstacles to automated data input, but IT-systems in production have changed a lot in the last 20 years. Skoogh et al. (2012) show how automated input data management can lead to time reduction and enhanced performance.

Several models and system architectures have been proposed to model the data exchange between physical and digital production systems. Those models are the foundation of the optimization and updating process of digital twins (Redelinghuys et al., 2020; Uhlemann et al., 2017). These works focus on the input side of the digital twin and updating, but do not discuss output validation and related automated update triggering in greater detail. Recent work of Goodall et al. (2019) presents a use case for data input in a remanufacturing facility.

2.3 Open research topic

Most of the existing approaches focus on automatic model generation. Some end up in a model translation, where the production system is modelled in a certain modelling style and then translated into an DES, which only decreases the modelling effort, if a model in the original modelling environment already exists (Terkaj and Urgo, 2015).

To tackle the problem of the need for initial modelling and because commercial simulation tools permit the easy and intuitive creation of simple models even for beginners, the presented approach chooses a different path: An existing model, which is manually modelled and implemented in a commercial simulation software, shall be enhanced by validation and update modules to turn it into a digital twin, which permits its use over the entire life time of the production system. The hypothesis is that the automated validation and updating can improve the initial models performance in terms of prediction accuracy.

3 Own Approach

The presented approach is explicitly aimed at simulation models of existing production systems, which shall be improved or controlled. It does not work for planning simulation models of production systems, which are not yet existing, since a comparison to reality and real data-based updates are impossible. Nevertheless, the approach can be used to transform planning simulation models into process accompanying simulation models during the building and commissioning phase of the production system.

3.1 Process flow

Production lines evolve over time and thus the input data needed for the simulation model, such as process times, availabilities, quality rates etc., change. Therefore, it is necessary to ensure that the digital twin always stays up-to-date and offers a close representation of reality in a given time period. The presented solution is composed of a two parts iterative process (Fig. 1): the validation and the automated updating procedure.

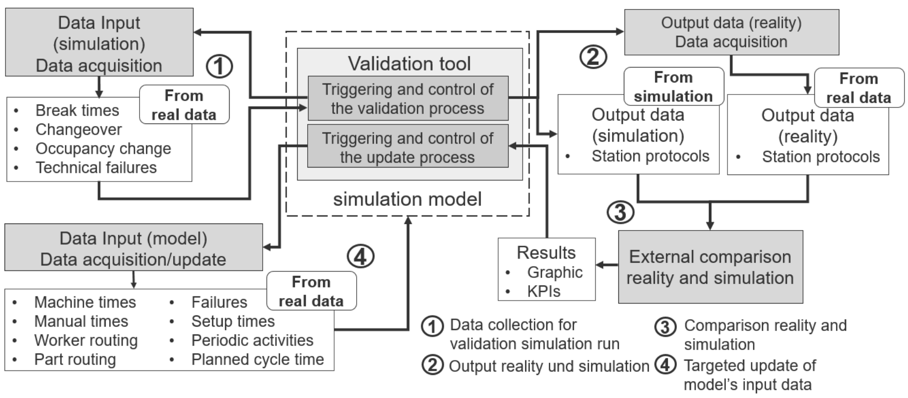

Figure 1: *Iterative process of validation and automated updating*

The simulation model itself is built and validated beforehand by simulation experts, following (VDI, 2014).

3.2 Validation

The objective of the validation is to automatically compare the simulation model with reality on different levels. The first step is to compare the output of the simulation model and reality by using carefully chosen key performance indicators (KPIs) and boundary values.

To evaluate the deviation of simulation runs to reality, the relative variation (see Eq. 1) and the NRMSE (Normalized Root Mean Square Error) (see Eq. 2) are used. The variation quantifies the final state of production of the studied period whereas the NRMSE quantifies the difference between reality and simulation during the course of the studied period.

$$Variation = \frac{\|N_{real} - N_{Sim}\|}{N_{real}} * 100 \qquad (1)$$

with N_{real} , N_{sim} being the total amount of produced part at the end of the studied period respectively in reality and in simulation.

$$NRMSE = \frac{1}{\bar{x}_{real}} * \sqrt{\frac{\sum_{i=1}^{N}(x_{real} - x_{sim})^2}{N_{real}}} \qquad (2)$$

with x_{real} and x_{sim} representing the total amount of produced parts at each point in time t_i of the studied period, respectively for reality and simulation.

If the model output values deviate from the real output less than a predefined degree, this means that the digital twin satisfies the expectations and represents the reality to a satisfactory extent. In the case that outputs do not match, input values of the digital twin have to be examined in order to differentiate between input parameters that are still up-to-date and obsolete ones. According to these analysis results, the automated updating will be triggered precisely for the relevant parameters.

3.3 Update

In order for the automated updating process to be efficient, two prerequisites have to be fulfilled. A digital twin where the most effective input parameters are characterized as well as a data pipeline between data sources and simulation system are indispensable. Furthermore, the automated updating process allows replacing outdated data.

Once the update is performed, a simulation run is realized and the validation process is repeated to check the validity of the updated model. The whole process is repeated until the output is within the boundaries or until the digital twin cannot be further improved. In this case feedback is given to the user that an appropriate level of closeness could not be reached automatically and a manual intervention is necessary. An important outcome of this iterative process is to choose an appropriate time period for the data acquisition, that consequently gives the best compromise between data meaningfulness and acquisition effort while satisfying the performance criteria of the digital twin.

4 Use Case

The described approach was developed in a research partnership between of the wbk Institute for Production Science at the Karlsruhe Institute of Technology (KIT) and the central department Connected Manufacturing of the Bosch Powertrain Solutions division with the goal to develop an agile production system. Its application and validation are also part of this joint research project.

4.1 Production system

The exemplary production system, for which the digital twin is implemented, assembles car engine components in high volume and is composed of two areas which are connected via a conveyor. The two areas are assembly and testing, each semi-automated, following the Chaku-Chaku principle. This means that the machines perform their processes mainly in an automated manner and the workers are primary required for loading and unloading of machines and transporting parts between them. The line produces various product types with differing material flows, processing times, etc. The number of workers in each area varies due to external factors as vacations, sick days, reduced customer demand, trainings, etc. This has to be considered in the validation of the model. Historic production data from various sources is stored in a central data lake, including process times, change over delays, errors, scrap rates, etc. The software “Tecnomatix Plant Simulation” by Siemens is used to implement the digital twin.

4.2 Implementation

Using the approach described above, a validation tool that enables the validation and automated updating process is implemented. Before running the simulation model, the validation has to gather information about the system status at each point of time of the validation period from existing information systems such as MES and ERP. This includes the number of workers, produced product types and exceptionally long downtimes (more than one hour), that appear very rarely. If the simulation run would not consider this information, its comparison to reality would not be meaningful. The information about the number of workers in the production system at a certain period in time is not stored in the data lake, but in a different IT-System which is not accessible and therefore has to be added manually.

A python script preprocesses the real and simulation output data and compares them automatically. In the use case the chosen characteristic KPIs are: the progression of produced parts over time, the variance of the hourly overall equipment effectiveness (OEE) as well as the total OEE within the analyzed time period. These KPIs give an overview over the systems performance and keep track of the behavior of the digital twin during the whole simulation run. The permitted deviation of each KPI is decided according to the company’s performance goals and the systems inherent fluctuation. In the use case the corresponding threshold of permitted deviation shall not exceed 3% for the variation and 5 for NRMSE.

If the validation process results in higher deviations, another Python script performs the automated update by directly accessing the IT systems and data warehouses to obtain the latest input data. The data pipeline is composed of SQL queries and then filtered and processed into exploitable update data for the Plant Simulation software.

5 Results

Three experiments were conducted on two different weeks of production. The first experiment validates an input data set and the model`s behaviour while the second and third experiment highlight the use of an automated targeted update to correct the input data and enable a better fitting of simulation with reality.

5.1 Automated validation

The first experiment was conducted for a production period of one week. To model the non-deterministic behaviour of the simulation, five simulation runs with different random seed values were conducted for each experiment with Plant Simulation Tecnomatix to get a statistical confidence of the results. Those five runs were considered sufficient as they well reflect the statistical repartition of the model while ensuring an acceptable optimized run time of the experiment. The automated validation compares each simulation run with reality and on the one hand returns figures (Figure 2) to help the user visualize the part's production during the production period. On the other hand, it generates key values to quantify the production systems behaviour (Table 1). Figure 2 shows a good fit of line output between the simulation runs and the reality for both the assembly and testing lines. This visual analysis is confirmed by the calculated key values from Table 1. The mean variation for both lines is under 3% and the mean NRMSE is under 5.

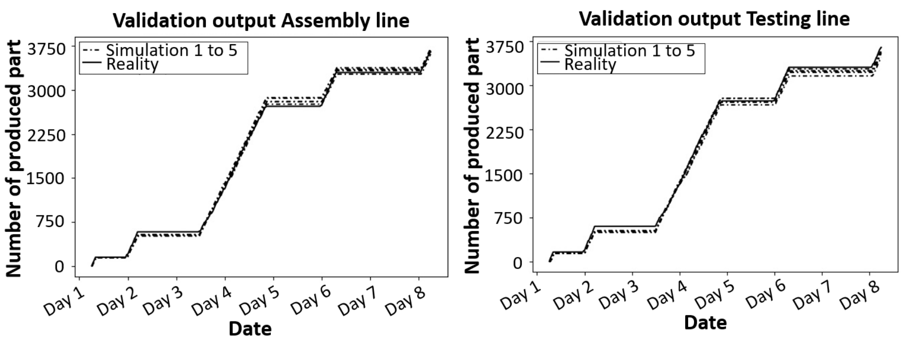

Figure 2: *Validation output of assembly and testing line*

Table 1: *Results of automatic validation*

Line	Produced parts reality	Mean produced parts simulation	Mean variation (%)	Mean NRMSE
Assembly	3650	3677	1.13	2.69
Testing	3639	3545	2.58	3.57

The fixed criteria from section 4.2 are therefore fulfilled and the input data is considered still up-to-date. The focus of the analysis lies on the number of produced parts since the OEE follows this number linearly.

5.2 Targeted update of the input parameters

For the second and third experiment, simulation and validation were conducted for another week of production. In the second experiment, the same input parameters as in section 5.1 were used. However, the obtained results before any update (Table 2, Figure 3) from the validation process exceeded the fixed threshold. Therefore, an update of input parameters is triggered. The first step of the update process is to determine which data must be replaced and if the line is partially or totally concerned by the update. The mean variation on both assembly and testing line are bigger than 3%, furthermore the NRMSE of the testing line is above 5. Consequently, both lines have to be updated.

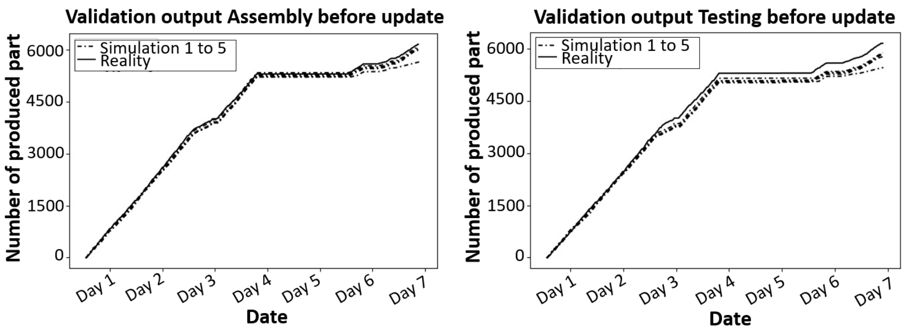

Figure 3: *Validation output before update on assembly and testing line*

Among the input data, it is possible to update the following parameters: Part routing, worker routing, failures, machine process times, manual process times, setup time and planned cycle times. Nonetheless, among those parameters few register notable variations during the chosen time period. In this paper, the focus was put on the machine process times, which encountered consequent variation over the studied week. After recalculating the probability density function of the machine process times from real data with a python script, the targeted update process compares the new calculated values with the old values for each machine. The machine process times are modelled by a normal distribution through mean and standard deviation. If the mean differs more than 0.15 seconds and the standard deviation more than 0.2, the old value is replaced with the new value. As mentioned above, in this use case the other input data did not change significantly and did not need any update.

Once the input parameters are updated, a third experiment with the newly calculated input data is conducted. Figure 4 depicts the output validation after the update for assembly and testing lines. Figure 4 shows improvement compared to Figure 3. The behaviour of the simulation is closer to reality and shows less variability. Those observations are verified through the key values in Table 2. For the assembly line, the mean variation of simulation went down from 3.03% to 0.92% and the mean NRMSE went from 2.99 to 1.77. For the testing line the mean variation went down from 6.32% to 3.2% and the NRMSE from 5.69 to 3.22. A net improvement is indeed realized. The behaviour of the assembly line is now completely validated whereas the testing line still has a mean variation barely above 3%. But the NRMSE has been improved and is now below 5. The machine process times could not be further improved for the testing line. In a next step, other parameters of the simulation models, i.e. availabilities, scrap rate, etc. should be updated. For these parameters an automated update process is not yet implemented.

The capability of the targeted update process was nonetheless proved but still needs further improvement particularly concerning the threshold values and the trigger conditions for the targeted update mechanism.

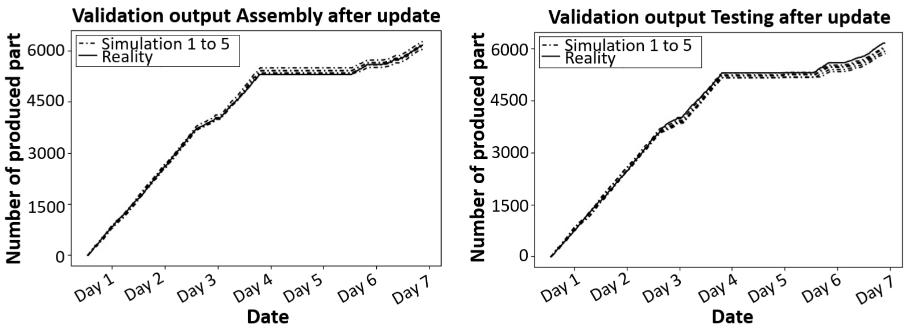

Figure 5: *Validation output after update on assembly and testing line*

Table 2: *Validation metrics before and after automatic update*

Experiment number	Line	Mean difference output (%)	Mean NRMSE output	Validation passed?
Before update	Assembly	3.03	2.99	No
After update	Assembly	0.92	1.77	Yes
Before update	Testing	6.32	5.69	No
After update	Testing	3.2	3.22	No

6 Conclusion and Outlook

Motivated by the ever-changing structure of modern production systems, an approach to enable simulation models to mirror these changes was developed. The approach contains a module for continuous validation which compares simulation KPIs to real historic KPIs. If a certain deviation threshold is surpassed, this module triggers an automated update module which changes the simulation model to better reflect reality. The application of this approach at a semi-automated production line of automotive components leads to a convergence of the simulation model to reality, turning it into a digital twin.

Further research has to be done to evaluate the behaviour of the digital twin in different scenarios of changes in the production system as well as its robustness to incomplete and/or biased data. Another line of research would be the extension of the available update mechanism of the digital twin. This could be combined with a thorough examination of the validation KPIs and their thresholds.

References

Bergmann, S.: Automatische Generierung adaptiver Modelle zur Simulation von Produktionssystemen. Illmenau, TU Illmenau, Dissertation, 2013.

Bergmann, S.; Straßburger, S.: Automatische Modellgenerierung – Stand, Klassifizierung und ein Anwendungsbeispiel. In: Mayer, G.; Pöge, C.; Spieckermann, S.; Wenzel, S. (Hrsg.): Ablaufsimulation in der Automobilindustrie. Berlin, Heidelberg: Springer Berlin Heidelberg 2020, pp. 333–347.

Goodall, P.; Sharpe, R.; West, A.: A Data-Driven Simulation to Support Remanufacturing Operations. Computers in Industry (2019) 105, pp. 48–60.

Kotiades, K.: Towards self-adaptive discrete event simulation (SADES). In: Anagnostou, A.; Hoad, K.; Kunc, M. (Hrsg.): Operational Research Society Simulation Workshop 2016 (SW16) 2016,

Kritzinger, W.; Karner, M.; Traar, G.; Henjes, J.; Sihn, W.: Digital Twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine 51 (2018) 11, pp. 1016–1022.

Mayer, G.; Pöge, C.; Spieckermann, S.; Wenzel, S. (Hrsg.): Ablaufsimulation in der Automobilindustrie. Berlin, Heidelberg: Springer Berlin Heidelberg 2020.

Redelinghuys, A.J.; Basson, A.H.; Kruger, K.: A six-layer architecture for the digital twin: a manufacturing case study implementation. Journal of Intelligent Manufacturing 31 (2020) 6, pp. 1383–1402.

Reinhardt, H.; Weber, M.; Putz, M.: A Survey on Automatic Model Generation for Material Flow Simulation in Discrete Manufacturing. Procedia CIRP 81 (2019) 2, pp. 121–126.

Robinson, N.; Perera, T.: Automated data collection for simulation? Simulation Practice and Theory 9 (2002) 6-8, pp. 349–364.

Skoogh, A.; Johansson, B.; Stahre, J.: Automated input data management: evaluation of a concept for reduced time consumption in discrete event simulation. SIMULATION 88 (2012) 11, pp. 1279–1293.

Splanemann, R.: Teilautomatische Generierung von Simulationsmodellen aus systemneutral definierten Unternehmensdaten. Bremen, Universität Bremen, Institut für Betriebstechnik und angewandte Arbeitswissenschaften, Dissertation, 1995.

Terkaj, W.; Urgo, M.: A Virtual Factory Data Model as a Support Tool for the Simulation of Manufacturing Systems. Procedia CIRP 28 (2015) 1, pp. 137–142.

Uhlemann, T.H.-J.; Lehmann, C.; Steinhilper, R.: The Digital Twin: Realizing the Cyber-Physical Production System for Industry 4.0. Procedia CIRP 61 (2017), pp. 335–340.

VDI: VDI-Richtlinie 3633: Simulation von Logistik-, Materialfluß- und Produktionssystemen, Berlin, Beuth, 2014,

Werner, S.; Weigert, G.: Process accompanying simulation - a general approach for the continuous optimization of manufacturing schedules in electronics production. In: Proceedings of the Winter Simulation Conference, San Diego, CA, USA, 8-11 Dec. 2002, 2002, pp. 1903–1908.