Comprehensive Validation Metrics and Precise Updating of Digital Twins of Production Systems

Leonard Overbeck[1*], Arthur Le Louarn[1], Oliver Brützel[1], Nicole Stricker[2], Gisela Lanza[1]

[1]wbk Institute of Production Science, Karlsruhe Institute of Technology, Kaiserstr. 12, 76131 Karlsruhe, Germany;
*leonard.overbeck@kit.edu

[2]Hochschule Aalen - Technik und Wirtschaft, Beethovenstr. 1, 73430 Aalen, Germany

SNE 32(3), 2022, 135-142, DOI: 10.11128/sne.32.tn.10613
Selected ASIM SPL 2021 Postconf. Publication: 2022-01-18;
Received English version: 2022-05-10; Accepted: 2022-06-03
SNE - Simulation Notes Europe, ARGESIM Publisher Vienna
ISSN Print 2305-9974, Online 2306-0271, www.sne-journal.org

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pace to respond to rapidly changing markets and business environments, simulation models of these systems become outdated earlier, reducing their useful time window.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.

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 [1]. Yet, in most companies simulation models of production systems are still built and used only in temporary projects [2]. 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 [2] 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 [3].

1 Literature Review

1.1 Model generation and maintenance

[4] 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, [5] 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.

[6] 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 [7] which uses the Core Manufacturing Simulation Data standard to create simulation models. [8] 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 [9].

1.2 Data input for simulation models

[10] 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. [11] shows 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 [12,13]. 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 [14] presents a use case for data input in a remanufacturing facility.

1.3 Model validation

It exists a broad literature concerning the validation of computer models ranging from general discussion of different statistical tests methods [15,16] over the assesment of the general forecast possibilities in particular domains [17] to finetuning of specific tests in certain domains, for example ecological modelling [18].

There are also various works on the challenge of simulation model validation [19].

A general practical guide for validation was published and applied to two discrete simulation models of manufacturing cells by [20].

1.4 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 [21].

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.

2 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.

2.1 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. The simulation model itself is built and validated beforehand by simulation experts, following [2].

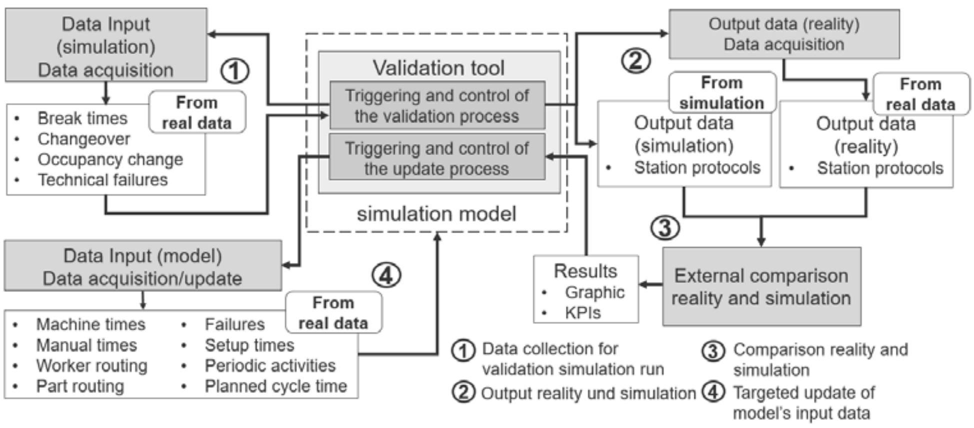

Figure 1: Iterative process of validation and automated updating.

2.2 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. Further steps are an in-depth analysis using regression analysis.

Deviance measures

To evaluate the deviation of simulation runs to reality, the relative error (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.

$$Relative\ error = \frac{\lVert N_{real} - N_{Sim} \rVert}{N_{real}} * 100 \tag{1}$$

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.

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

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

Another possibility to measure the prediction error of the simulation model is Theil's U2 (see Eq.3) which becomes 0 for a perfect prediction and 1 if equal to the naïve prediction [22]. When using Theils U2 it is important to know that big prediction errors have a greater influence on the metric [23].

$$U_2 = \frac{\sqrt{\frac{1}{T-h} \sum_{t=h+1}^{T} (x_{sim,t} - x_{real,t})^2}}{\sqrt{\frac{1}{T-h} \sum_{t=h+1}^{T} (x_{real,t})^2}} \tag{3}$$

Regression analysis

To get an even better understanding of the behaviour of digital twin and reality, it is also helpful to look at the linear regression fit of actual versus predicted values [24]. One important parameter to measure the difference between the simulation and real system using this regression approach is R² [25].

$$R^2 = \frac{\sum_{i=1}^{n} (\widehat{x_i^{real}} - \overline{x^{real}})^2}{\sum_{i=1}^{n} (x_i^{real} - \overline{x^{real}})^2} \tag{3}$$

The regression fit can also be described using the intercept (a) defined in Eq.4 and the slope b (Eq.5) of the regression line. For a perfect fit the slope would be 1 and the intercept would be 0.

$$a = \overline{x^{real}} - b * \overline{x^{sim}} \tag{4}$$

$$b = \frac{\sum_{i=1}^{n} (x_i^{sim} - \overline{x^{sim}}) * (x_i^{real} - \overline{x^{real}})}{\sum_{i=1}^{n} (x_i^{sim} - \overline{x^{sim}})^2} \tag{5}$$

Consequences of validation

If the model output values deviate from the real output less than a predefined degree, it means 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.

2.3 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 reach 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.

3 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.

3.1 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, machine failures, scrap rates, etc. The software “Tecnomatix Plant Simulation” by Siemens is used to implement the digital twin.

3.2 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 manufacturing execution systems (MES) and enterprise resource planing (ERP) systems. 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 variation 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 behaviour of the digital twin during the whole simulation run. The permitted deviation of each KPI is decided on accordingly 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 relative error 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.

4 Results

Four experiments were conducted on three different weeks of production. The first experiment validates an input data set and the model`s behaviour with the basic KPIs and their static thresholds, while the second experiment provides an in-depth analysis of the behaviour using more advanced regression and statistical KPIs. The third and fourth experiments highlight the use of an automated targeted update to correct the input data and enable a better fitting of simulation with reality.

4.1 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 graphs (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.

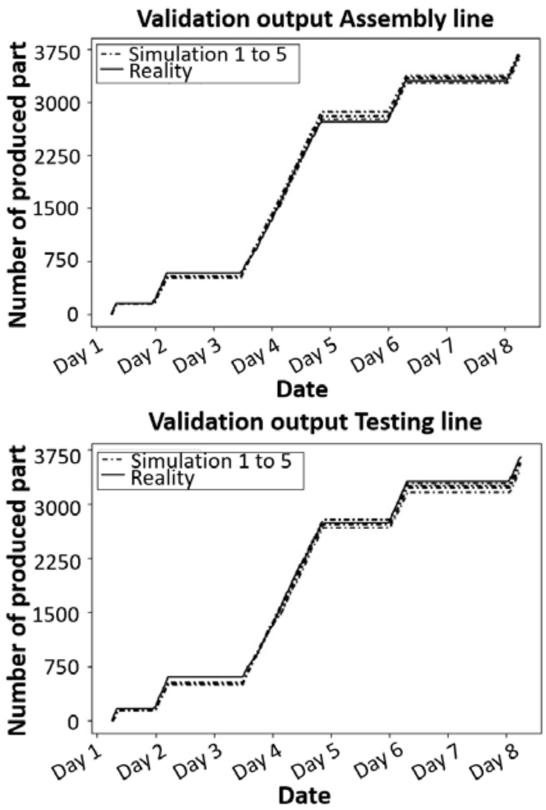

Figure 2: Validation output of assembly and testing line – experiment 1.

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

Table 1: Results of automatic validation.

The fixed criteria from Section 3.2 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.

4.2 In-depth analysis

The second experiment focuses on an in-depth validation of another week which has a different production plan with different product variants and different production breaks. The visual analysis in Figure 3 already indicates a good fit between the curves of reality and the simulation runs. When we are now looking at the mean deviation of overall output, we see that it is 2.0% and 1.64%, which is lower than the defined thresholds, therefore the validation is positive according to this indicator. The NRMSE is 4.55 for assembly and 4.82 for testing, which also indicates a good fit of the model given the threshold of 5.

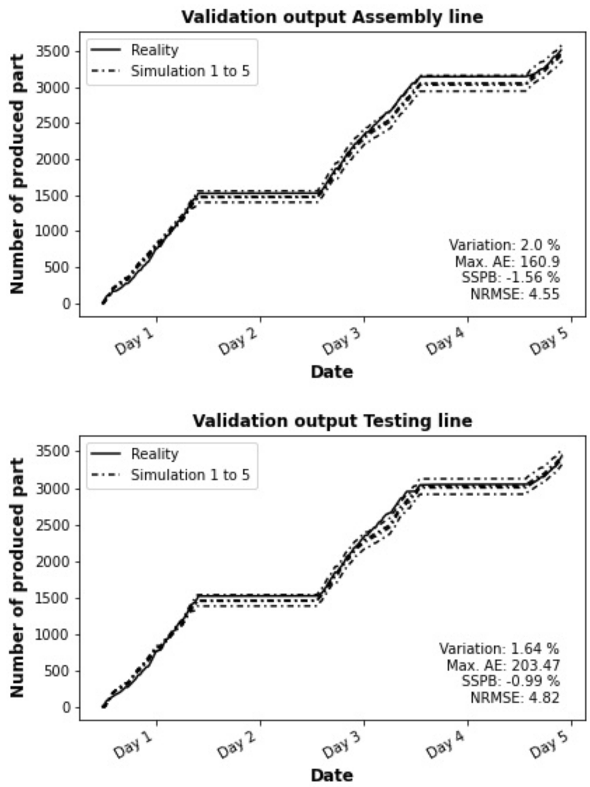

Figure 3: Validation output of assembly and testing line – experiment 2.

Figure 4 shows the regression analysis of experiment 2 including the related parameters for assembly and testing. R^2 reaches with 0.86 and 0.9 quite high values which expresses a good fit of the simulation model to reality as well. The slope is in both cases very close to 1 which is another indicator that the model in general provides a good estimation for the behaviour of the real system. The intercepts are 0.56 and 0.96 which is also a good value considering the absolute scale of the axis ranges from 0 to 100.

Therefore, the regression analysis underlines the positive validation result of the subjective visual impression and the basic deviation analysis. Theil's U_2 is 0.52 and 0.54 respectively which also indicates a good fit.

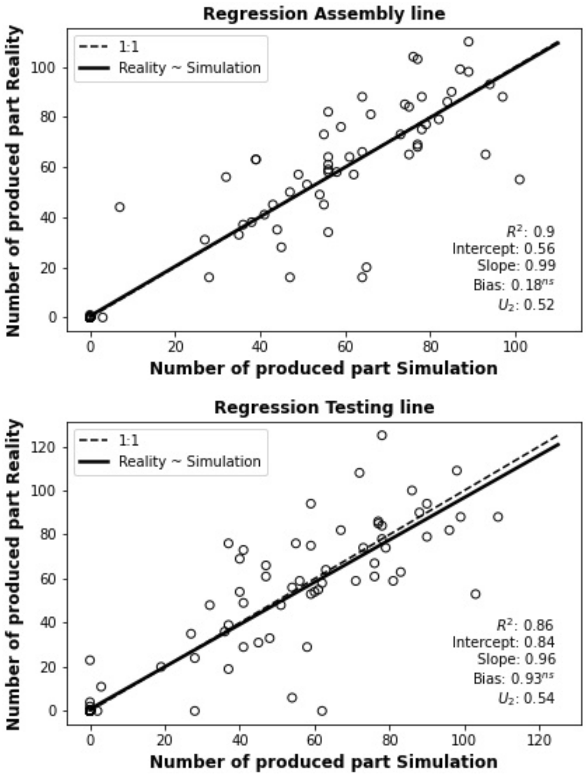

Figure 4: Regression analysis of testing and assembly line - experiment 2.

4.3 Targeted update of the input parameters

For the third and fourth experiment, simulation and validation were conducted for another week of production. In the third experiment, the same input parameters as in Section 5.1 were used. However, the obtained results before any update (Table 2, Figure 5) 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 relative error 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.

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.

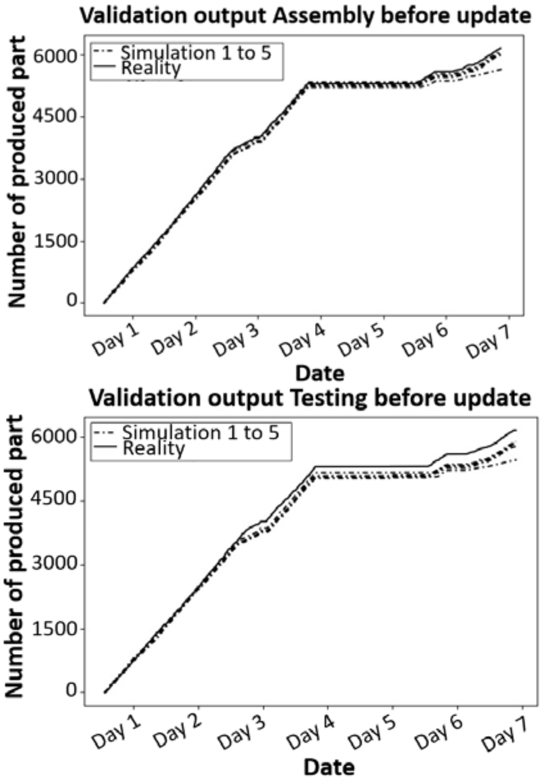

Figure 5: Validation output before update on assembly and testing line – experiment 3.

Nonetheless, among those parameters few register notable deviation 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 fourth experiment with the newly calculated input data is conducted. Figure 6 depicts the output validation after the update for assembly and testing lines. Figure 6 shows improvement compared to Figure 5. 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 relative error of simulation went down from 3.03% to 0.92% and the mean NRMSE went from 2.99 to 1.77.

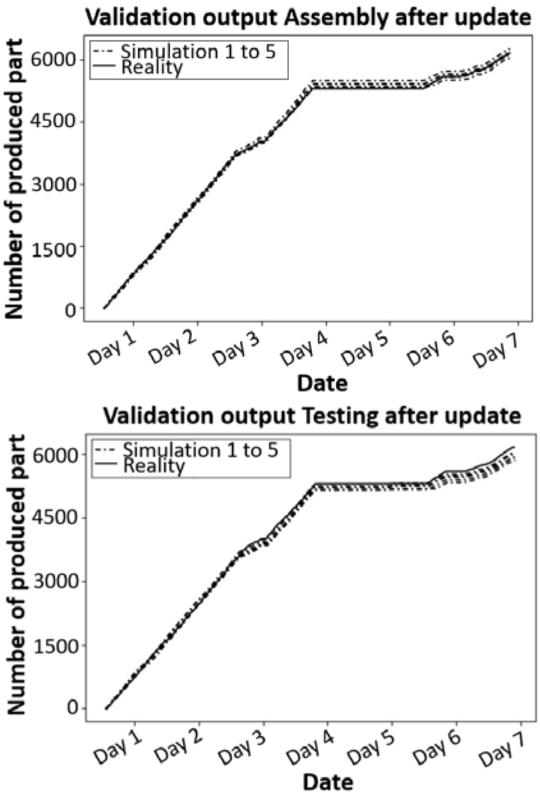

Figure 6: Validation output after update on assembly and testing line – experiment 4.

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.

Experiment	Line	Mean difference (%)	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

Table 2: Validation metrics before and after automatic update.

For the testing line the mean relative error 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 relative error 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.

5 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. Various metrics to measure the deviation of the simulation to reality for this validation module where discussed and implemented ranging from simple deviation KPIs to more complex statistical and regression values. 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. This includes the further study of the behaviour of the various reality metrics introduced to this paper. 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

[1] Mayer G, Pöge C, Spieckermann S, et al., editors. Ablaufsimulation in der Automobilindustrie. Berlin, Heidelberg: Springer Berlin Heidelberg; 2020.

[2] VDI. VDI-Richtlinie 3633: Simulation von Logistik-, Materialfluß-und Produktionssystemen. Berlin; 2014.

[3] Kritzinger W, Karner M, Traar G, et al. Digital Twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine. 2018;51: p. 1016–1022.

[4] Splanemann R. Teilautomatische Generierung von Simulationsmodellen aus systemneutral definierten Unternehmensdaten [Dissertation]. Bremen: Universität Bremen; 1995.

[5] 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. Proceedings; 8-11 Dec. 2002; San Diego, CA, USA; 8-11 Dec. 2002; p. 1903–1908.

[6] Bergmann S, Straßburger S. Automatische Modellgenerierung – Stand, Klassifizierung und ein Anwendungsbeispiel. In: Mayer G, Pöge C, Spieckermann S, et al., editors Ablaufsimulation in der Automobilindustrie; 2020; p. 333–347.

[7] Bergmann S. Automatische Generierung adaptiver Modelle zur Simulation von Produktionssystemen [Dissertation]. Ilmenau: TU Ilmenau; 2013.

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

[9] Reinhardt H, Weber M, Putz M. A Survey on Automatic Model Generation for Material Flow Simulation in Discrete Manufacturing. Procedia CIRP. 2019;81: p. 121–126.

[10] Robertson N, Perera T. Automated data collection for simulation? Simulation Practice and Theory. 2002;9:349–364.

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

[12] Redelinghuys AJH, Basson AH, Kruger K. A six-layer architecture for the digital twin: a manufacturing case study implementation. Journal of Intelligent Manufacturing. 2020;31: p. 1383–1402.

[13] Uhlemann TH-J, Schock C, Lehmann C, et al. The Digital Twin: Demonstrating the Potential of Real Time Data Acquisition in Production Systems. Procedia Manufacturing. 2017;9: p. 113–120.

[14] Goodall P, Sharpe R, West A. A Data-Driven Simulation to Support Remanufacturing Operations. Computers in Industry. 2019: p. 48–60.

[15] Mayer DG, Butler DG. Statistical validation. Ecological Modelling. 1993;68: p. 21–32. DOI 10.1016/0304-3800(93)90105-2.

[16] Buturac G. Measurement of Economic Forecast Accuracy: A Systematic Overview of the Empirical Literature. Journal of Risk and Financial Management. 2022;15: p. 1. https://www.mdpi.com/1415108.

[17] Teng PS. Validation of computer models of plant disease epidemics: A review of philosophy and methodology/: Zuverlässigkeit von Computermodellen für Epidemien von Pflanzenkrankheiten: Ein Überblick über Grundgedanken und Methodik. Zeitschrift für Pflanzenkrankheiten und Pflanzenschutz / Journal of Plant. 1981;88: p. 49–63. https://www.jstor.org/stable/43214777.

[18] Piñeiro G, Perelman S, Guerschman JP, et al. How to evaluate models: Observed vs. predicted or predicted vs. observed? Ecological Modelling. 2008;216: p. 316–322. DOI 10.1016/j.ecolmodel.2008.05.006

[19] Sargent RG. An introductory tutorial on verification and validation of simulation models. In: 2015 Winter Simulation Conference (WSC). Proceedings; 06.12.2015 - 09.12.2015; Huntington Beach, CA, USA; 06.12.2015 - 09.12.2015; p. 1729–1740.

[20] Leal F, Costa, Rafael Florêncio da Silva, Montevechi JAB, et al. A practical guide for operational validation of discrete simulation models. Pesqui. Oper. 2011;31: p. 57–77. DOI 10.1590/s0101-74382011000100005.

[21] Terkaj W, Urgo M. A Virtual Factory Data Model as a Support Tool for the Simulation of Manufacturing Systems. Procedia CIRP. 2015;28: p. 137–142.

[22] Theil H. Economic Forecasts and Policy. The Review of Economics and Statistics. 1961;43: p. 310. http://dx.doi.org/10.2307/1927302.

[23] Barrot C. Prognosegütemaße. In: Albers S, Klapper D, Konradt U, editors Methodik der empirischen Forschung; 2007; p. 417–430.

[24] Analla M. Model validation through the linear regression fit to actual versus predicted values. Agricultural Systems. 1998;57: p. 115–119. DOI 10.1016/s0308-521x(97)00073-5.

[25] Smith EP, Rose KA. Model goodness-of-fit analysis using regression and related techniques. Ecological Modelling. 1995;77: p. 49–64. DOI 10.1016/0304-3800(93)e0074-d.