Validation of Data-Driven Reliability Models in Manufacturing - Work in Progress

Jonas Friederich[1*], Sanja Lazarova-Molnar[2,1]

[1]Mærsk Mc-Kinney Møller Institute, University of Southern Denmark, 5230 Odense, Denmark; [*]*jofr@mmmi.sdu.dk*

[2]Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

SNE 34(3), 2024, 163-166, DOI: 10.11128/sne.34.sn.10698
Selected EUROSIM 2023 Proceedings Publication: 2024-06-18
Received Revised Improved: 2024-07-15; Accepted: 2024-07-24
SNE - Simulation Notes Europe, ARGESIM Publisher Vienna
ISSN Print 2305-9974, Online 2306-0271, www.sne-journal.org

Abstract. Reliability modeling enables deriving reliability measurements and illustrating relevant fault-dependencies in manufacturing systems. Data-driven reliability modeling uses data generated in systems to either automate or at least support extraction of reliability models. To use these extracted models for decision support, we need to ensure models' validity. In this extended abstract, we discuss our initial approach for validating data-driven reliability models. The challenge with validating data-driven models lies in the fact that these models are continuously generated and updated, implying that we need a new or updated validation approach to enable an ongoing validation of these models. The upside is that the systems of interest generate large amounts of data, which can significantly support the quantitative validation processes. Additionally, we briefly address the implications that could result from our proposed approach.

Introduction

Advancements in manufacturing technology have led to the generation and collection of vast amounts of data that are stored within information systems, such as Manufacturing Execution Systems (MES) and Supervisory Control and Data Acquisition (SCADA) systems.

This data, generated from equipment control systems, such as Programmable Logic Controllers (PLCs) or collected from sensors monitoring equipment state, can be used to support decisions. On the flip side, modern manufacturing systems have become increasingly complex, which complicates systems' maintenance and identification of possible vulnerabilities that affect their reliabilities.

To this end, reliability modeling includes a number of techniques to assist with this. Conventional reliability modeling, however, relies significantly on expert knowledge of the system under study, which can become a bottleneck as systems become more complex and experts sparse [1]. Moreover, manufacturing systems are often subject to frequent modifications that can quickly make such conventional reliability models obsolete when systems' topologies change [2]. Thus, there is a need to dynamically generate accurate reliability models for manufacturing systems based on data to address the challenges described above and to ensure optimal exploitation of reliability models in the shopfloors [3, 4]. This is what we term as data-driven reliability modeling [5].

Validation is necessary to enable the use of the data-driven reliability models to support decisions. Conventionally, model validation is carried out by a subject matter expert after the model has been constructed. However, the automatic generation of models in data-driven reliability modeling requires an automated approach to continuous validation [6].

The vast amounts of data generated by advanced manufacturing systems can be utilized to address these challenges and enhance the accuracy and statistical significance of validation outcomes.

Here, we introduce our approach for validating data-driven reliability models (Section 1) and discuss data requirements, as well as other implications arising from our proposed approach (Section 2).

1 Validation of Data-driven Reliability Models

Figure 1 outlines the approach that we propose for validation of data-driven reliability models for manufacturing systems and how it is embedded in the general process of data-driven reliability assessment (DDRA).

The general process of DDRA consists of the following phases:

1. Definition, generation, collection and preprocessing of relevant data,
2. reliability model extraction,
3. validation of the extracted model,
4. simulation and calculation of systems' reliability measures and
5. presentation of results in a dashboard to support decision-making [5].

To validate extracted reliability models, we follow the typical two steps:

1. ensuring face validity, and
2. quantitative validation [7].

Face validity is used to describe a subjective judgment of experts whether the model and/or its behavior accurately reflect the real-world system being modeled. Face validity can be assessed by reviewing the model structure, inputs, and outputs and comparing them to the real-world system, as well as by conducting a qualitative evaluation of the results.

Quantitative validation compares data from the real system with data generated from the simulation model, with the goal of applying statistical hypothesis testing to determine the similarity of the simulation model and the real system with respect to predefined performance measures [8].

The availability of data streamed from information systems, such as MES or SCADA, as well as from sensors, offers an opportunity to automate and enhance the quantitative validation, which is what we aim to explore.

Quantitative validation can be performed through either input-output transformations, or streaming input data. Input-output transformations compare output data from the real system with the output data from the simulation model, without utilizing real data for the input variables.

However, in the case of advanced manufacturing systems, the generation of vast amounts of data presents an opportunity to feed a simulation model with sufficient data, such that the necessary number of replications can be performed with it to yield validation results with the required level of statistical significance. Since we derive reliability models from data, we can also use the same data to quantitatively validate extracted reliability models. Furthermore, the continuous recording of data in the physical system enables continuous validation in addition to periodic and on-demand validation.

If the validity of a reliability model is not refuted, it implies that the model can be used to support relevant decisions. For example, we can evaluate the impact that different resources have on the overall reliability of the system.

If the validity of a reliability model is refuted, we must regenerate (i.e., repeating the first two phases of DDRA, as described earlier) or, in case of minor issues, recalibrate the model.

2 Discussion

In this last section we describe the data requirements for validating data-driven reliability models. We then highlight various research directions for utilizing validation, such as model calibration, as well as considerations for scheduling validation and determining the appropriate timing for generating new models.

To enable validation of data-driven reliability models, it is important to gather and have access to all the necessary data that are required for the validation process. This includes, for example, event logs that capture relevant events related to material flow in a system and a state log that captures state changes in the system's production resources.

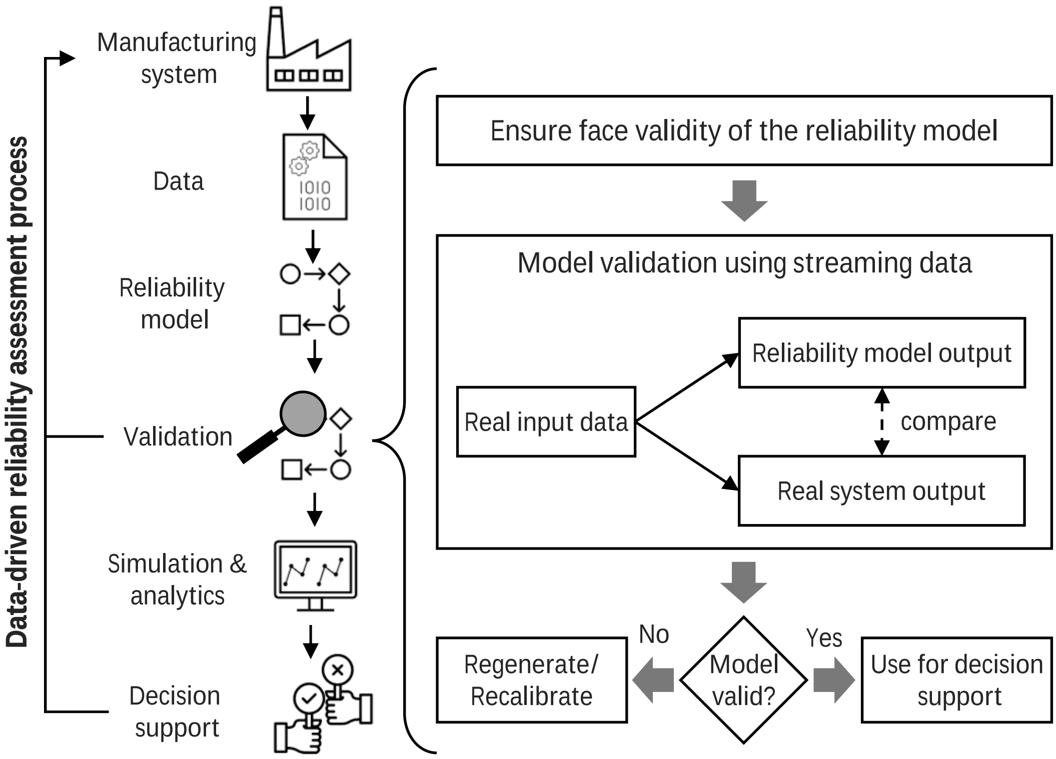

Figure 1: Proposed approach for validation of data-driven reliability models.

In the previous section, we described that the same data used to generate the model can also be used for its validation. Performance indicators such as throughput (i.e., the number of completed orders per time unit) and downtime of the system and/or its resources can be used for validation.

The information needed to calculate these indicators is readily available in the event log capturing material flow and the state log tracking resource state changes.

Validation is a vital component of data-driven simulation modeling in general, as validation can be used to calibrate extracted models [8].

One approach to quantitatively validate and calibrate an extracted reliability model is through the use of reinforcement learning (RL). RL can be utilized to optimize the models' parameters to better approximate the behavior of the manufacturing system. When the simulation model's output deviates from the system's actual output, an RL-agent can make adjustments to the model parameters.

After a each simulation run, the agent either receives a positive or negative reward based on the simulation results. RL can also be used to aid calibrating the reliability model to optimize towards a given performance indicator, such as throughput.

For example, if the agent identifies that increasing buffer sizes or reducing failure rates of production resources would lead to increased throughput, it can trigger a reconfiguration of the manufacturing system.

Another important aspect we need to consider is the scheduling of model validation. This can be either time-based or trigger-based. For example, in critical systems the reliability model should be validated in real time to ensure continuous robustness of the model. In less critical systems, validation can be scheduled, for example, once a day or once a week. A new validation run can be triggered by the handling of a new production batch, change of shift, change of resources/equipment, or simply, if the output of the model seems incorrect.

Clearly, a new reliability model must be generated and thus validated once there is a change in topology/-configuration of the physical entity (e.g., introduction of redundancy, change of production routes, change of maintenance policy).

Furthermore, the integration of new data sources can be used to enrich the data-driven reliability model, which in turn requires validation of the enriched model.

With this extended abstract, we aim to open a discussion and to stimulate research on validation of data-driven reliability models for manufacturing systems. This is especially relevant in the emerging context of digital twins. In future, we plan to test our proposed approach in a case study.

References

[1] Friederich J, Francis DP, Lazarova-Molnar S, Mohamed N. A framework for data-driven digital twins of smart manufacturing systems. *Computers in Industry*. 2022; 136:103586.

[2] Lugaresi G, Matta A. Automated Manufacturing System Discovery and Digital Twin Generation. *Journal of Manufacturing Systems*. 2021;59:51–66.

[3] Lazarova-Molnar S, Li X. Deriving simulation models from data: steps of simulation studies revisited. In: *2019 Winter Simulation Conference (WSC)*. IEEE. 2019; pp. 2771–2782.

[4] Lazarova-Molnar S, Mohamed N. Reliability assessment in the context of industry 4.0: data as a game changer. *Procedia Computer Science*. 2019;151:691–698.

[5] Friederich J, Lazarova-Molnar S. Towards Data-Driven Reliability Modeling for Cyber-Physical Production Systems. *Procedia Computer Science*. 2021; 184C:589–596.

[6] Friederich J, Lazarova-Molnar S. A Framework for Validating Data-Driven Discrete-Event Simulation Models of Cyber-Physical Production Systems. In: *Proceedings of the 2023 Winter Simulation Conference*. San Antonio, USA: IEEE. 2023; .

[7] Sargent RG. Verification and Validation of Simulation Models. In: *Proceedings of the 2010 Winter Simulation Conference*. 2010; pp. 166–183.

[8] Hua EY, Lazarova-Molnar S. Validation of Digital Twins: Challenges and Opportunities. In: *Proceedings of the 2022 Winter Simulation Conference*. 2022; .