Proceedings of the 2022 Winter Simulation Conference
B. Feng, G. Pedrielli, Y. Peng, S. Shashaani, E. Song, C.G. Corlu, L.H. Lee, E.P. Chew, T. Roeder, and P. Lendermann, eds.

VALIDATION OF DIGITAL TWINS: CHALLENGES AND OPPORTUNITIES

Edward Y. Hua

The MITRE Corporation
7525 Colshir Drive
McLean, VA 22102, USA

Sanja Lazarova-Molnar

Institute of Applied Informatics and Formal Description Methods
Karlsruhe Institute of Technology, Keiserstr. 89
Karlsruhe, 76133, GERMANY

Deena P. Francis

DTU Compute
Technical University of Denmark,
Kongens Lyngby, 2800, DENMARK

ABSTRACT

Digital twins enjoy increasing interest in a diverse array of industrial sectors, such as manufacturing, healthcare, urban planning, etc. Their usefulness depends on the robustness of the corresponding digital twin models; however, validation of the model, as a mean to ensure models' robustness, is a difficult problem. Moreover, traditional validation approaches need to undergo significant transformation to be made applicable to digital twins. To the best of our knowledge, there has not been a systematic treatment of validating digital twin models. This paper identifies several challenges facing the model validation within digital twins. Furthermore, we propose an initial framework to define basic rules of digital twin model validation and introduce a systematic approach to validation that seamlessly combines expert knowledge and data gathered from available Internet of Things (IoT) devices.

1 INTRODUCTION

Simulation has long been utilized in academia, industry, and government as a powerful tool to model real-life systems, permitting analysts and other stakeholders to study system behaviors and provide insights into systems operations. In recent years, more and more systems, from a wide spectrum of fields, have been employing simulation for various applications, including, but not limited to, theory testing (Mizuno et al. 2018), scientific modeling (Jones and Luyten 2018; Awais et al. 2020), system performance analysis (Sarda and Digalwar 2018; Entrialgo et al. 2021), and operator training (Hernández et al. 2021).

As systems become more sophisticated and intricate, so do the demands for their models to reach a higher degree of realism to ensure that they accurately mirror the real-time, dynamic state of systems. A traditional simulation model is disadvantaged to meet such modeling challenges in several respects (Gómez-Romero and Molina-Solana 2021; Lazarova-Molnar and Li 2019). It is typically built with historical system data (or, in the absence of actual data, with synthetic data). A fixed set of pre-defined parameters governs the simulation throughout its runtime. Parts of the model may be abstracted to achieve a faster execution time, but at the expense of reduced model fidelity. Furthermore, simulation results are usually available only after the end of the run, and it is often a time-consuming process to analyze the data and produce data products to inform the relevant system stakeholders of the model performance. If the model is poorly designed and developed, its output results would not accurately reflect the true state of the system. The traditional simulation model, burdened with these shortcomings, is ill-suited to support the modeling requirements of many of today's systems (Jeong et al. 2020).

A digital twin is a digital representation of a physical object, process, or service (Fuller et al. 2020). The digital representation consists of attributes and properties that uniquely characterize the physical entity. It replicates the physical entity to a higher degree of detail resolution than a traditional simulation model. A well-built digital twin model allows the user to continuously monitor the performance of the physical entity, detects faults/anomalies in real-time, analyzes the data and suggests a remedial solution, and presents the solution back to the physical entity. Connecting the physical and digital counterparts are two data-flow conduits, one going from the physical to the digital, and the other from the digital to the physical. The data-flow conduits can be fully automated, or they can be intervened with human operators.

Since the early 2000s, when the term was first introduced, digital twins have been applied to an increasingly diverse array of industries, such as advanced manufacturing, healthcare, automotive, and urban planning. Its continuing evolution, from conceptualization to practical applications, is in large part due to the explosive growth of Internet-of-Things (IoT), in which small, inexpensive, and energy-efficient wireless sensors are mass-produced and deployed to a wide range of industrial and commercial products. These sensors continuously collect data on the physical entity's state of health and operation and transmit the data to the system's digital twin that uses it to perform tasks such as model updating, monitoring, diagnosis, and fault detection. If the digital twin uses the received data to discover any operational deficiency of the physical entity, it sends back a solution to be implemented by the physical entity to resolve the deficiency.

The usefulness of a digital twin largely rests on its robustness, captured by the digital twin's ability to closely mirror the current state of its physical counterpart. This is largely reflected by the quality of the underlying simulation model, typically assessed through a validation process. Model validation is the procedure of ensuring that the model the observed performance of the actual system closely matches the "synthetic" performance computed from the model output. In the context of digital twins, this means that the digital twin model needs to be validated to mimic the actual state of the physical counterpart to a high degree of accuracy, given the simulation goals of the digital twin.

While an expansive body of literature on model validation exists, there is relatively little published work that specifically addresses the problem space of digital twin model validation. Fully realizing the potential that digital twins have to offer requires a high-resolution (i.e., sufficiently-detailed), high-fidelity (i.e., operationally faithful to the physical system) model, a timely transfer of collected data from the physical entity to its digital twin, a real-time analytical and resolution capability, and an effective mechanism to send the analytical output back to the physical entity (Fuller et al. 2020). All these requirements place a high premium on accurately validating the digital twin model. In this paper we propose a general framework of digital twin model validation. It shall serve as a guidepost to incentivize further discussions and interest in this critical problem.

The remainder of this paper is organized as follows. Section 2 provides relevant background and related work on model validation. In Section 3, we identify the challenges of a successful digital twin model validation and propose a general framework in addressing these challenges. Section 4 discusses how applying this validation framework would open new areas of research. Section 5 concludes the paper.

2 BACKGROUND AND RELATED WORK

2.1 Model Validation

Developing a digital twin model typically involves several steps, including data collection, data validation, knowledge extraction, model development, and model validation (Friederich et al. 2021). One key step that is largely under-addressed in the literature is model validation.

Model validation refers to the confirmation that the model of the physical entity conforms to the expected system performance requirements. This is different from model verification, where the objective is to ensure that the model has been built correctly according to the requirements or specifications. Model validation is typically performed as a two-step process: face validity and quantitative validation (Sargent 1991; Martis 2006). A number of approaches have been utilized for establishing face validity of a simulation model, which is in general largely subjective. Some of the more popular are animation and Turing test (Sargent 1999; Liu and Yang 2005; Tsioptsias et al. 2016; Jantke et al. 1997), as the goal is to check the simulation model's realism and appropriateness when compared to the actual simulation goal. The second step is more quantitative and objective in that it is comparing data from the real system with data generated from the simulation model (White and Sinclair 2004), with the aim to utilize statistical hypothesis testing methods and determine the similarity in the output of the simulation model and the real system with respect to the predefined performance measures. Such an approach falls into the category of input validation (Nelson 2010). In our paper, we assume the logic validation approach (Nelson 2010), where the same input data generated by the real system is fed into the digital twin model for computations. Logic validation also serves as the basis of trace-driven simulation (Lugaresi et al. 2019; Lugaresi and Matta 2018).

Presently, there does not exist a methodology or systematic framework for digital-twin validation. The lack of a systematic approach to digital twin approach could lead to varying levels of model fidelity and robustness, which would impact the effectiveness of the digital twin model. If a human operator is part of the process, an ill-validated model would directly affect the operator's ability to make informed decisions in response to an anomalous phenomenon in the physical entity.

2.2 Digital Twins

Since the term "digital twin" was first coined by Michael Grieves of the Florida Institute of Technology in 2002, it has been given a number of definitions to fit its wide adoption (Nath 2021). In this paper, we treat a digital twin as a digital representation of a system, process, or service that exists in the physical world. It captures the characteristics and system-dynamic behaviors of the physical entity with high model resolution. Throughout its life cycle, data concerning the health and performance of the physical entity is continuously collected by sensors attached to said entity, and in real-time transmitted to its digital twin to reinforce the robustness of the model and monitor the performance of its physical counterpart. If the data indicates a performance degradation, anomaly, or fault in the physical entity, it can perform analysis to find the cause of the issue and suggest the best solution that can be fed back to its physical counterpart to implement.

Digital twins evolved from simulation, at a time when it leveraged the rapid growth of Internet-of-Things (IoT). It differs from simulation in some significant ways. Firstly, simulation uses historical data to drive its execution; digital twin, on the other hand, necessitates the input of real-time data from the physical entity to show its worth. Secondly, simulation is typically driven with a fixed set of parameters that governs its execution throughout a simulation run; digital twin, by virtue of its use of real-time data, could mature over time, and is not constricted by the set of parameters it initially started with. Thirdly, as it uses historical data to drive its progression, simulation tends to produce results that may not accurately inform the current state of the system under test, whereas the digital twin is much more responsive to the dynamics of system behavior.

It is worth noting that digital twins and traditional simulation play an equally important role in the current and future Modeling & Simulation (M&S). Decisions on applying either traditional simulation methods or digital twins must depend on what the objective of the system under test is. For example, traditional simulation is often sufficient to evaluate the performance of a wireless communications network (e.g., throughput, latency, packet completion rate, etc.), as it does not necessarily require detailed modeling of all the communications devices of the network. Digital twins, on the other hand, may be needed to ensure the smooth operations of said network in real-time, since a high-resolution digital representation of the actual network at the core of the agent-based model is essential in taking into considerations all possible factors that could cause network operations to fail. Judiciously choosing the right approach during the planning stage ensures that excessive resources will not be spent on achieving the task objective.

2.3 Related Work on Digital Twin Model Validation

The quality of the information a digital twin sends back to its physical counterpart hinges upon the robustness and accuracy of the digital twin model. Digital twin model validation remains a challenging and inadequately investigated problem. There is scant published literature that touches upon this subject. In (PivotPointTechnologyCorporation 2021), one approach that is suggested is a reasonable litmus test, akin to the Turing Test for AI. It assumes a System Tester (ST) that runs a robust system Verification & Validation (V&V) Test Suite on both the physical system twin (PT) and its digital twin (DT) but cannot reliably distinguish between with a probability greater than 80%, the DT is declared a bona fide digital twin of the PT. This approach may be difficult to set up, since it is not easy to acquire a System Tester with a V&V test suite. Wright and Davidson considered trust in the model, along with trust in the data and data-updating procedure, to be one of the key areas to ensure confidence in reports made by the digital twin (Wright and Davidson 2020). (Argota Sánchez-Vaquerizo 2021) identified several challenges when attempting to validate a class of large-scale digital twins for urban traffic using empirical data, such as the unreliability of the source data, trade-off between efficiency and realism, and lack of quantitative assessment with empirical measurements.

We identify the following as general validation strategies from the current literature:

- **Manual / visual inspection**: It refers to the process of experts visually inspecting the digital twin and verifying its correctness according to established standards. The problems with visual inspection are that they can be tedious, error-prone and there is also the possibility of overlooking certain aspects. This type of validation is often carried out offline.
- **Property testing**: It is a method of formally testing various important properties of digital twin. This is comparatively better than manual / visual inspection as testing can be automated with software, and even digital twins themselves that perform validation (Löcklin et al. 2020). The problems with property testing are that it can be tedious and time consuming if the number of properties being tested is very large, which is often the case of digital twins. This is also often carried out offline.
- **Model based testing**: It is a method that interacts with the digital twin either online or offline and tests certain properties of it to confirm its correctness (Utting and Legeard 2010). A specific kind of testing within model-based testing is referred to as the Input-Output Conformance testing (IOCO) (Tretmans 1996). This method uses a specification containing the valid inputs and a list of valid outputs. If the model outputs anything outside of those specified within the specification, given a set of valid inputs, it flags this behavior as needing further investigation. (Khan et al. 2018) used an offline model-based testing approach based on IOCO for digital twins within the context of legacy systems.
- **Machine learning based validation approaches**: This approach is used in scenarios where the goal is to validate certain functionalities of the digital twin. Examples of such specialized scenarios include fault detection. (Farhat et al. 2021) used a feature selection and a supervised classification approach to perform fault severity detection. The digital twin in this instance was built to simulate ball bearing fault detection. Machine learning approach was specifically used for predicting the type of fault using the data generated by the digital twin. The validation aspect of this work consists of performing a correlation analysis on extracted features from actual signals and simulated signals. The discrepancy between measured and simulated signals serves as the basis for updating the parameters of the digital twin.

3 VALIDATION OF DIGITAL TWIN MODELS

3.1 Open Problems

(Wright and Davidson 2020) points out that, due to the dynamic nature of system behavior, a validated model can provide a snapshot of the physical entity's behavior only at a specific moment, which may become obsolete at a later time. Validation of a digital twin model should be a continuous process that needs to be performed periodically (or on demand) throughout its lifecycle. Such time-variant, system-dynamic behavior of the physical entity leads to a number of open problems that impede the robust validation of a digital twin model. We discuss each open problem in the following subsections.

3.1.1 Modeling Realism

Part of the digital twin's appeal is its high level of modeling realism, underpinned by model resolution and fidelity. A digital twin model is different from a traditional simulation model in that, while certain aspects of the simulation model can be abstracted out to achieve some efficiencies (e.g., ease of model build-up, improved speed of simulation run, analytical tractability, etc.), a trade-off of achieving such efficiencies is the degraded realism of the model. A digital twin model, on the other hand, must replicate its physical counterpart with a high degree of accuracy to function as a mirror representation of its physical counterpart. Supporting the necessary resolution and fidelity of the digital twin model requires detailed characterizations of the physical entity. When these characterizations change over time, they must be re-acquired and re-synthesized. Maintaining model realism that accurately reflects the physical entity's current state in a timely manner directly impacts the model's ability to remain fully cognizant of system operations.

3.1.2 Data Uncertainty

Since a digital twin model is data-driven, uncertainty in the input data naturally arises when it is being sent to the digital twin model. Data uncertainty can originate from a large number of sources, from the mechanisms (e.g., sensors) that take the measurements, to range tolerance in specification elements of the product requirements documentation. (Ríos et al. 2020) have pointed out that there often exists a lack of knowledge about the uncertainty of data captured in the physical entity. Uncertainty, in general scientific computing, can be classified into two types: Epistemic and aleatoric (Kreye et al. 2011; Roy and Oberkampf 2011). Epistemic uncertainty arises from the lack of knowledge (insufficient assumptions, missing data, inaccurate models, etc.), and aleatoric uncertainty is inherent in the non-deterministic variability in the physical process. As components of the physical entity are not independent of each other, data uncertainty in one component is likely propagated and compounded. As result, captured measurements could deviate significantly from the true values of the system performance. Furthermore, much of this uncertainty is expressed only in qualitative terms. A proper treatment of data uncertainty, e.g., as a stochastic process, is needed to aid in the periodic model validation quantitatively.

3.1.3 System Dynamics

Even if a digital twin model is properly validated at a particular time epoch, due to the often-dynamic nature of the physical entity, by the time the digital twin sends its analytical output to the physical entity, the state of the physical entity has changed such that the latest validation becomes obsolete. This problem may be particularly acute when there is a long delay in either the data transfer from the physical entity to its digital twin (which could include a series of steps such as data collection, data transfer, and data reduction), or in the analysis and assessment of the physical entity's state. The most common approach to addressing this problem is proactive data sampling, in which the IoT sensors on the physical entity periodically transmit data to its digital twin, which performs computations to monitor the system's operational performance and, if necessary, detect and resolve any perceived anomalies. This approach could potentially quickly detect an anomaly or drastic changes in the system. However, it may require the support of substantial computational resources to store offloaded data and perform needed analysis.

3.1.4 Use Case Alignment

During the life cycle of a physical entity, its operational objective may frequently change at different stages, even when all assets of the physical entity do not. A new operational objective defines a new use case for the physical entity and may require a re-balancing and re-appropriation of the assets to support it. The digital twin model that mirrors the physical entity now must be judiciously aligned with each new use case. At the present, the only way to re-align the use case is to stop the operations and manually configure the digital twin model. How to ensure a timely reconfiguration of the model to mirror the physical entity without pausing its operations remains an open problem.

3.1.5 Reporting Invalid Models

Another problem that remains largely unaddressed in the literature is the lack of an effective mechanism for the digital twin to report an invalid model. Two events may cause a model to be declared invalid. Firstly, when an anomaly occurs in the physical entity and impacts its operations, the performance degradation is detected by its digital twin, which, by virtue of its definition, means that the digital twin model now deviates from the current state of the physical entity, hence the model becoming invalid. Secondly, the physical entity proactively makes an operational change for a new objective (e.g., to re-appropriate the resources for a new task).

In both events, recipients of the invalid model report should include, at a minimum, the analytical component of the model (either offline or online), the physical entity (e.g., querying for additional input data), and the stakeholders (for situational awareness and/or possible human intervention). However, the goals of the reporting are different in these events. In the first event, the goal is for the model to devise a solution to mitigate the adverse effects and send it to the physical entity for implementation. In the second event, the goal for reporting to the model's analytical component is to develop a new set of performance metrics for evaluating the new objective; to the physical entity, the reporting goal should effect potential changes in the types of data the sensors will collect and pass on to the model that can be used to derive the new metrics.

3.2 Proposed Digital Twin Validation Framework/Strategy

To respond to the challenges imposed by the new developments within the Industry 4.0 development (as pointed out in Section 3.1) and utilize the benefits that the new and emerging associated technologies bring along, we summarized our findings towards a framework for validation of digital twins. The main aspects that the framework is built upon are: availability of continuously streaming data from IoT devices, digital twin extracted through process mining approaches in combination with event detection and distribution fitting, and expert-in-the-loop for all expert knowledge necessity, e.g., relevant events and data streams specification. Process mining is a family of techniques that aim to discover process flows of a system, based on structured logs (Van der Aalst 2013). The digital twin validation framework focuses on the second step of the 2-step validation process, i.e., the step that is concerned with the data-driven and quantitative validation of the simulation model, i.e., the digital twin. The framework is inspired by the validation using historical data.

We assume that there will be some degree of human expert knowledge as part of the framework, but its entry points will be fully specified. One of the main aspects where expert knowledge is necessary is the specification of simulation goals and relevant data streams and events, which is also very relevant for the validation. Simulation goals also yield the performance measures for the digital twin that will need to be provided as output. The digital twin will then also need to be run using the same input data as the real system, in order to pair-compare the outputs.

To illustrate and contextualize our proposed framework, we use the illustration from our previous work on data-driven digital twins (Friederich et al. 2022), shown in Figure 1 with the validation part highlighted. Here, we expand and revise on the validation part of the digital twin. The validation also utilizes the collected data, but it also needs the parts of the model that have been extracted, and that includes the model's structure/topology as well as its parameters in terms of probabilities, constants and probability distributions.

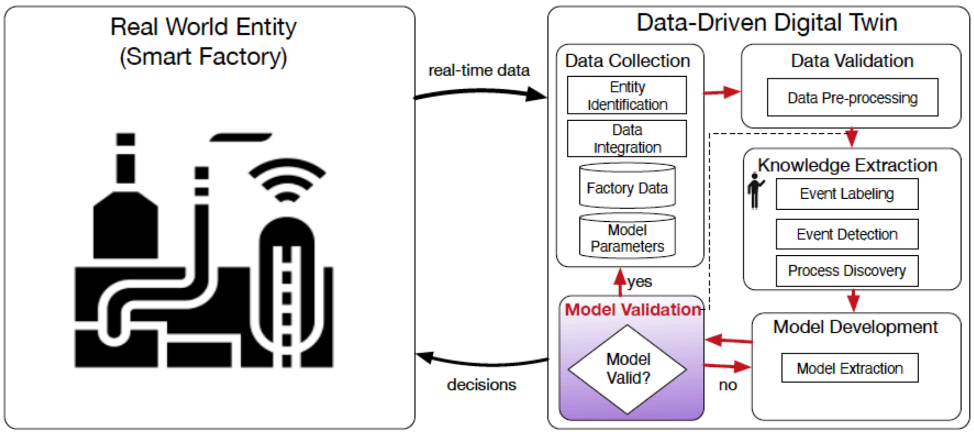

Figure 1: Framework for data-driven digital twins (Friederich et al. 2022).

The validation component (illustrated in Figure 2) would utilize data gathered from the smart factory, after the data has been validated. We exemplify our validation model through a simple factory line in which we collect input data, i.e., data on starting times of product orders, as well as machine status data (e.g., begin and end of an operation and failing or getting repaired), and output data on production key performance indicators, e.g., products completed per hour, reliability of the system, etc. (Lazarova-Molnar and Mohamed 2019). When we consider smart factories, equipped with IoT devices and for which there are digital twins available, all data, both input and output, is easily available so validation using either input-output transformations or historical input data is possible at any point in time. The exact method, however, depends on the available quantity of historical input data, i.e., whether we have sufficient data to feed into our simulation model, such that it can be used to run the necessary number of replications to yield validation results of the required level of statistical significance.

Therefore, in the case when an initial model is already available, usually developed through factory planning and derived from various machine specifications, validation can be performed using input-output transformations, in which only the output data from the real system is compared with the output data from the simulation model, without utilizing real data for the input random variables. This framework takes advantage of the availability of streaming data and, in combination with clearly identified input by experts, performs validation either periodically or on-demand. The data-driven models also benefit from this validation process as the outputs of the validation process are directly fed into the model extraction processes. In a way, this type of validation is also necessary when one considers data-driven model extraction, as it can support the processes of parameter calibration and model extraction.

This framework takes advantage of the availability of streaming data and, in combination with clearly identified input by experts, performs validation either periodically or on-demand. The data-driven models also benefit from this validation process as the outputs of the validation process are directly fed into the model extraction processes. In a way, this type of validation is also necessary when one considers data-driven model extraction, as it can support the processes of parameter calibration and model extraction.

Das ist das was Christian vorgeschlagen hat

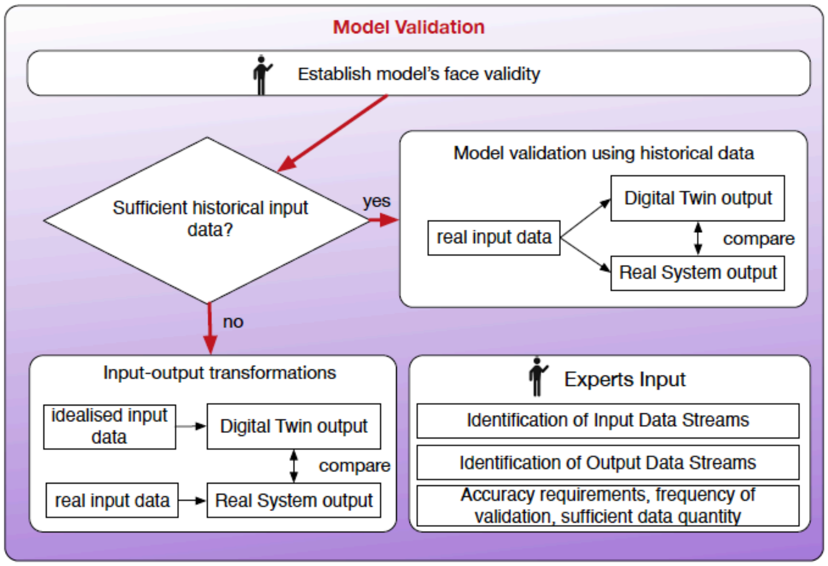

Figure 2: Validation component as part of the framework for data-driven digital twins.

4 IMPLICATIONS/DISCUSSION

We now discuss the implications of our study on validations of digital twins and its associated challenges. In Section 4.1 we describe the anticipated challenges related to the digital twin model validation. In Section 4.2 we discuss the opportunities that future related research may unlock.

4.1 Challenges

We describe in more detail, the challenges that may present with the validation of digital twin models.

4.1.1 Uncertainty and Sensitivity Analysis

Any data collected from a physical system inherently contains uncertainty that must be judiciously treated to achieve a degree of robustness in model validation. Uncertainty analysis permits the modeler to quantify such data uncertainty. Given that the apparatus of a digital twin model consists of a multitude of sensors attached to the physical entity to collect data in real-time, uncertainty analysis need be conducted with data from diverse data sources. Different data types may require different algorithms to quantify data uncertainty. If not carefully planned, these algorithms could incur substantial computational costs that contribute to a prolonged model validation process.

Model Averaging?

In model validation, uncertainty analysis often is a necessary step that precedes sensitivity analysis. As one needs to manage a large number of parameters when monitoring the digital twin model, it may become a necessity to identify a subset of them that are more impactful than the others. The sensitivity analysis may be utilized to do this. Applying sensitivity analysis and identifying a smaller, more manageable set of parameters would have the following advantages: 1) Reduce computational complexity when validating the model. 2) Reduce the time to validate the model. 3) Identify important connections between observations, model inputs, and predictions or forecasts, leading to the development of better models (Hill et al. 2016; Hill and Tiedeman 2006). Sensitivity analysis is almost always performed by running the model a large number of times, e.g., a sampling-based approach (Helton et al. 2006). This may not be feasible in the context of digital twin models, where real-time validation is often required. How to improve the runtime of sensitivity analysis remains an open problem. To the best of our knowledge, many digital twin models do not go through a rigorous process of uncertainty and sensitivity analyses.

4.1.2 Model Validation of System-of-Systems

A digital twin model that represents a System-of-Systems (SoS) substantially increases the difficulty of model validation. Each individual component in the SoS represents a system that supports the overall functionality. This presents a unique challenge to validating the digital twin model: Should validation be performed over the entire SoS, or for each constituent system? While one may argue that it is the overall performance of the SoS that matters, concentrating only on the digital twin model of the entire SoS risks missing the gradual deviation of some constituent system model's performance from its physical counterpart. One possible consequence is failure to identify the root cause in SoS performance degradation.

How to perform robust validation of the SoS remains a key challenge. To address this challenge, we may need to have a two-layer approach to perform validation, one at the SoS level, and the other at the constituent system level. This inevitably increases the complexity of the validation process and may affect the expediency of validation. How to balance the validation complexity, robustness, and timeliness remains a work in progress.

4.1.3 Combining Expert Knowledge and Collected Data

A simulation model should always utilize all information that we have about the real system. In that sense, some of the knowledge may stem from experts and might complement what we extract from data. Digital twins and corresponding validation approaches need to systematically provide mechanisms to combine expert knowledge with data.

There has been some notable efforts in tackling this issue (Niloofar and Lazarova-Molnar 2021), but the problem is far from a solution, and the solution may need to be context-dependent. Thus, we need approaches to formalize the expert knowledge and its combination with the data in a seamless manner, such that both can be integrated to yield a simulation model.

4.2 New Opportunities Presented by the Ongoing Data Collection

There are two aspects of availability of resources that bring about opportunities for validation of digital twins. The first aspect is the data, and the second is the developments in modelling approaches.

Opportunities may arise out of the availability of data. First, huge data sets provide opportunities to build models with good generalizability. This is of course strongly tied to the task at hand. The model(s) are built with a very specific objective in mind, for example fault detection. In this application, access to data such as sensor readings can be used to adapt the model to learn a wide range of variations in faults. The intuition is that the more data of a specific kind of fault that we have, the more the model can learn the structure of data which can be useful for discriminatory purposes and finally for decision making.

Second, large data sets enable the testing a suite of digital-twin properties using different varieties of data at different granularities. In addition to the large quantities of data that is available, different kinds of data are also available. By this we refer to the heterogeneous (multi-modal) nature of data. There are various possible ways of validating a digital twin model and testing various aspects of the model can be aided by the availability of varied kinds of data. For instance, in the case of fault detection, audio signals, thermal and imaging sensors are some of the data that can potentially be used for testing whether the model is robust enough to capture faults from sound, thermal and visual signals. This can also be extended to other applications where a suite of properties can be tested by means of the heterogeneous data that are collected.

Additionally, opportunities may also arise out of the developments in modelling approaches. The first is transfer learning. This technique is beneficial in dynamic environments such as the ones modelled by a digital twin. This is because transfer learning allows the use of already trained models from other domains to be re-used, which gets rid of the problem of learning from scratch each time new data is generated. (Maschler et al. 2021) explores the possibility of using transfer learning in the context of a cyber physical system through the various phases of manufacturing. Transfer learning can also be used to support the validation of a digital twin by utilizing the "knowledge" learned from other phases or types of validation.

Another opportunity in new modeling approach development is reinforcement learning. The dynamic nature of the validation process of a digital twin makes it possible to exploit methods suited to such environments. Reinforcement learning is one such approach that can be applied to the validation process. This method has been used in the context of digital twin development (Cronrath et al. 2019), process planning (Müller-Zhang et al. 2020). Digital twin validation can benefit from the use of appropriate reinforcement learning techniques.

5 CONCLUSION

Model validation is a critical step towards ensuring the digital twin model accurately mirrors the current state of the physical entity throughout its lifecycle. While researchers of various knowledge domains have implemented validation strategies in their respective digital-twin design prototypes, by and large it has not been rigorously treated in a way that would permeate its effectiveness throughout the digital twin lifecycle. Relatively little published work on this problem exists in the literature.

In this paper, we identified five open problems concerning digital twin model validation: modeling realism, data uncertainty, system dynamics, use-case alignment, and reporting invalid modes. If left unaddressed, they could impede the further adoption of digital twins. We then proposed a digital-twin model validation framework as a roadmap towards addressing these problems. We also identified three domains where further study is warranted in the context of digital twins: Uncertainty and Sensitivity Analysis, model validation of System-of-Systems, and combining expert knowledge and collected data. Future research could open up opportunities where Big Data becomes an indispensable tool to support robust model validation, and techniques such as transfer learning and reinforcement learning can be employed to develop novel modeling approaches.

ACKNOWLEDGMENTS

The authors wish to express their sincere gratitude to Dr. Rob Wittman of the MITRE Corporation for his insightful comments in preparing this paper.

Dr. Hua's affiliation with The MITRE Corporation is provided for identification purposes only and is not intended to convey or imply MITRE's concurrence with, or support for, the positions, opinions, or viewpoints expressed by the author. This contribution has been approved for Public Release, Distribution Unlimited. Case Number 22-1452. The paper is a joint academic effort by the three authors. For the MITRE contribution, the copyright remains until transferred to the conference proceedings.

REFERENCES

Argota Sánchez-Vaquerizo, J. 2021. "Getting Real: The Challenge of Building and Validating a Large-Scale Digital Twin of Barcelona's Traffic with Empirical Data". *International Journal of Geo-information* 11(1):24-51.

Awais, M., F. S. Alshammari, S. Ullah, M. A. Khan, and S. Islam. 2020. "Modeling and Simulation of the Novel Coronavirus in Caputo Derivative". *Results in Physics* 19:103588.

Cronrath, C., A. R. Aderiani, and B. Lennartson. 2019. "Enhancing Digital Twins through Reinforcement Learning". In *2019 IEEE 15th International Conference on Automation Science and Engineering*, edited by N. Amato, T. Asfour, Y. Choi, N. Y. Chong, H. Ding, D. Lee, C. C. Lerma, J. Li, E. Marchand, D. Popa, D. Song, Y. Sun, and P. Valdastri, 293-298. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Entrialgo, J., M. García, J. L. Díaz, J. García, and D. F. García. 2021. "Modelling and Simulation for Cost Optimization and Performance Analysis of Transactional Applications in Hybrid Clouds". *Simulation Modelling Practice and Theory* 109:102311.

Farhat, M. H., X. Chiementin, F. Chaari, F. Bolaers, and M. Haddar. 2021. "Digital Twin-Driven Machine Learning: Ball Bearings Fault Severity Classification". *Measurement Science and Technology* 32(4):044006.

Friederich, J., D. P. Francis, S. Lazarova-Molnar, and N. Mohamed. 2022. "A Framework for Data-Driven Digital Twins of Smart Manufacturing Systems". *Computers in Industry* 136:103586.

Friederich, J., S. C. Jepsen, S. Lazarova-Molnar, and T. Worm. 2021. "Requirements for Data-Driven Reliability Modeling and Simulation of Smart Manufacturing Systems." In *2021 Winter Simulation Conference,* edited by M. Loper, C. Szabo, S. Kim, B. Feng, K. Smith, S. Masoud and Z. Zheng*,* 1-12. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Fuller, A., Z. Fan, C. Day, and C. Barlow. 2020. "Digital Twin: Enabling Technologies, Challenges and Open Research". *IEEE Access* 8:108952-108971.

Gómez-Romero, J., and M. Molina-Solana. 2021. "Towards Data-Driven Simulation Models for Building Energy Management". In *International Conference on Computational Science*, edited by M. Paszynski, D. Kranzlmuller, V. V. Krzhizhanovskaya, J. J. Dongarra, and P. M. A. Sloot, 401-407. New York: Springer Publishing Company.

Helton, J. C., J. D. Johnson, C. J. Sallaberry, and C. B. Storlie. 2006. "Survey of Sampling-Based Methods for Uncertainty and Sensitivity Analysis". *Reliability Engineering & System Safety* 91(10-11):1175-1209.

Hernández, E., M. Camacho, C. Leal-Costa, M. Ruzafa-Martínez, A. J. Ramos-Morcillo, E. Cazorla, and J. L. Díaz-Agea. 2021. "Does Multidisciplinary Team Simulation-Based Training Improve Obstetric Emergencies Skills?" *Healthcare* 9(2):170-183.

Hill, M. C., D. Kavetski, M. Clark, M. Ye, M. Arabi, D. Lu, L. Foglia, and S. Mehl. 2016. "Practical Use of Computationally Frugal Model Analysis Methods". *Groundwater* 54(2):159-170.

Hill, M. C., and C. R. Tiedeman. 2006. *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty.* Hoboken: John Wiley & Sons.

Jantke, K. P., R. Knauf, and T. Abel. 1997. "The Turing Test Approach to Validation". In *15th International Joint Conference on Artificial Intelligence,* August 23rd-29th, Nagoya, Japan, 35-45.

Jeong, Y., A. Singh, M. Zafarzadeh, M. Wiktorsson, and J. Baalsrud Hauge. 2020. "Data-Driven Manufacturing Simulation: Towards a CPS-Based Approach". In *SPS 2020*, edited by K. Safsten and F. Elgh, 586-596. Amsterdam: IOS Press.

Jones, J. W., and J. C. Luyten. 2018. "Simulation of Biological Processes". In *Agricultural Systems Modeling and Simulation*, edited by R. M. Peart and W. D. Shoup, 18-61. Boca Raton, Florida: CRC Press.

Khan, A., M. Dahl, P. Falkman, and M. Fabian. 2018. "Digital Twin for Legacy Systems: Simulation Model Testing and Validation." In *2018 IEEE 14th International Conference on Automation Science and Engineering,* August 20th-24th, Munich, Germany, 421-426.

Kreye, M. E., Y. M. Goh, and L. B. Newnes. 2011. "Manifestation of Uncertainty-A Classification". In *Proceedings of the 18th International Conference on Engineering Design*, edited by S.J. Culley, B.J. Hicks, T.C. McAloone, T.J. Howard, and W. Chen, 96-107. Copenhagen: Design Society.

Lazarova-Molnar, S., and X. Li. 2019. "Deriving Simulation Models from Data: Steps of Simulation Studies Revisited." In *2019 Winter Simulation Conference,* edited by N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 2771-2782. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lazarova-Molnar, S., and N. Mohamed. 2019. "Reliability Assessment in the Context of Industry 4.0: Data as a Game Changer". *Procedia Computer Science* 151:691-698.

Liu, F., and M. Yang. 2005. "Validation of System Models". In *IEEE International Conference on Mechatronics and Automation,* edited by J. Gu and P.X. Liu, 1721-1725. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Löcklin, A., M. Müller, T. Jung, N. Jazdi, D. White, and M. Weyrich. 2020. "Digital Twin for Verification and Validation of Industrial Automation Systems–a Survey". In *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation*, September 8th-11th, Vienna, Austria, 851-858.

Lugaresi, G., G. Aglio, F. Folgheraiter, and A. Matta. 2019. "Real-Time Validation of Digital Models for Manufacturing Systems: A Novel Signal-Processing-Based Approach." In *2019 IEEE 15th International Conference on Automation Science and Engineering,* August 22nd-26th, Vancouver, Canada, 450-455.

Lugaresi, G., and A. Matta. 2018. "Real-Time Simulation in Manufacturing Systems: Challenges and Research Directions." In *2018 Winter Simulation Conference,* edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 3319-3330. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Martis, M. S. 2006. "Validation of Simulation Based Models: A Theoretical Outlook". *Electronic Journal of Business Research Methods* 4(1): 39-46.

Maschler, B., D. Braun, N. Jazdi, and M. Weyrich. 2021. "Transfer Learning as an Enabler of the Intelligent Digital Twin". *Procedia CIRP* 100:127-132.

Mizuno, Y., Z. Younsi, C. M. Fromm, O. Porth, M. De Laurentis, H. Olivares, H. Falcke, M. Kramer, and L. Rezzolla. 2018. "The Current Ability to Test Theories of Gravity with Black Hole Shadows". *Nature Astronomy* 2(7):585-590.

Müller-Zhang, Z., P. O. Antonino, and T. Kuhn. 2020. "Dynamic Process Planning Using Digital Twins and Reinforcement Learning". In *IEEE International Conference on Emerging Technologies and Factory Automation*, September 8th-11th, Vienna, Austria, 1757-1764.

Nath, S. V., P. v. Schalkwyk, D. Isaacs. 2021. *Building Industrial Digital Twins: Design, Develop, and Deploy Digital Twin Solutions for Real-World Industries Using Azure Digital Twins.* Birmingham: Packt Publishing, Ltd.

Nelson, B. L. 2010. *Stochastic Modeling: Analysis & Simulation.* Chelmsford, USA: Courier Corporation.

Niloofar, P., and S. Lazarova-Molnar. 2021. " Fusion of Data and Expert Knowledge for Fault Tree Reliability Analysis of Cyber-Physical Systems." In *5th International Conference on System Reliability and Safety,* November 24th-26th, Palermo, Italy, 92-97.

PivotPointTechnologyCorporation. 2021. How to Verify & Validate a Digital Twin? https://digitalengineeringgroup.com/faq/how-to-v-and-v-digital-twin.html, accessed 1st May.

Ríos, J., G. Staudter, M. Weber, R. Anderl, and A. Bernard. 2020. "Uncertainty of Data and the Digital Twin: A Review". *International Journal of Product Lifecycle Management* 12(4):329-358.

Roy, C. J., and W. L. Oberkampf. 2011. "A Comprehensive Framework for Verification, Validation, and Uncertainty Quantification in Scientific Computing". *Computer Methods in Applied Mechanics and Engineering* 200(25-28):2131-2144.

Sarda, A., and A. K. Digalwar. 2018. "Performance Analysis of Vehicle Assembly Line Using Discrete Event Simulation Modelling". *International Journal of Business Excellence* 14(2):240-255.

Sargent, R. G. 1991. "Simulation Model Verification and Validation." In *Proceedings of 1991 Winter Simulation Conference*, edited by B. L. Nelson, W. D. Kelton, and G. M. Clark, 37-47. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

———. 1999. "Validation and Verification of Simulation Models". In *Proceedings of 1999 Winter Simulation Conference*, edited by P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 39-48. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Tretmans, J. 1996. "Test Generation with Inputs, Outputs and Repetitive Quiescence". *Software-Concepts and Tools* 17(3):103-120.

Tsioptsias, N., A. Tako, and S. Robinson. 2016. "Model Validation and Testing in Simulation: A Literature Review." In *5th Student Conference on Operational Research,* edited by B. Hardy, A. Qazi, and S. Ravizza, 1-11. Leibnitz, Germany: Dagstuhl Publishing Company.

Utting, M., and B. Legeard. 2010. *Practical Model-Based Testing: A Tools Approach.* Burlington: Morgan Kaufmann Publishers

Van der Aalst, W. M. 2013. "Process Mining in the Large: A Tutorial". In *3rd European Business Intelligence 2013 Summer School*, July 7th to 12th, Dagstuhl Castle, Germany, 33-76.

White, A. S., and R. Sinclair. 2004. "Quantitative Validation Techniques a Database.(I). Simple Examples". *Simulation Modelling Practice and Theory* 12(6):451-473.

Wright, L., and S. Davidson. 2020. "How to Tell the Difference between a Model and a Digital Twin". *Advanced Modeling and Simulation in Engineering Sciences* 7(1):1-13.

AUTHOR BIOGRAPHIES

EDWARD Y. HUA is a Lead Modeling & Simulation Engineer at the MITRE Corporation. He received his Ph.D. from the Department of Electrical and Computer Engineering, Cornell University. He is a member of INFORMS. His research interests include digital twins, simulation and AI, and wireless network analysis. His email address is ehua@mitre.org. His website is https://www.linkedin.com/in/edward-hua-ph-d-1296025/. His ORC ID is 0000-0002-6294-3122.

SANJA LAZAROVA-MOLNAR is a Professor at the Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology. She is also a Professor at the University of Southern Denmark, where she leads the research group Modelling, Simulation and Data Analytics. She is a Senior Member of The Institute of Electrical and Elctronics Engineering (IEEE), and currently serving as Director-at-Large on the Board of Directors of The Society for Modeling & Simulation International (SCS). Furthermore, she is Chair of IEEE Denmark and Vice-Chair of IEEE Denmark Women in Engineering Affinity Group. Her email address is sanja.lazarova-molnar@kit.edu.

DEENA P. FRANCIS is a Postdoctoral researcher at Cognitive Systems, DTU Compute, Technical University of Denmark. Her areas of interest include machine learning, in particular, time series analysis and Gaussian Processes. Her email address is dfra@dtu.dk. Her website is https://deenafrancis.github.io/.