Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

DERIVING SIMULATION MODELS FROM DATA: STEPS OF SIMULATION STUDIES REVISITED

Sanja Lazarova-Molnar

Mærsk Mc-Kinney Møller Institute
University of Southern Denmark
Campusvej 55
Odense, 5230, DENMARK

Xueping Li

Department of Industrial and Systems Engineering
University of Tennessee
851 Neyland Drive
Knoxville, TN 37996, USA

ABSTRACT

Simulation is typically a result of a tremendous amount of work performed by experts in various fields, usually in computer science, mathematics and the corresponding application area. The currently uncomplicated accessibility to data provides a significant opportunity to reduce the requirements for expert knowledge in some aspects, or at least to only utilize expert knowledge to supplement and validate data-derived models, or, vice versa, use collected data to confirm and validate existing expert knowledge. In this paper we explore the idea of derivation of simulation models from data. We, furthermore, survey and summarize related existing efforts for the most popular simulation paradigms, identifying benefits, opportunities and challenges, as well as discuss the ways in which the traditional simulation study processes are impacted.

1 INTRODUCTION

Simulation is the process of studying systems' behaviors through reproducing on computer the manners in which these systems operate in the real world (often referred to as *computer simulation*). Typical goal of simulation studies is providing decision support on how to enhance a given system of interest with respect to certain performance measures. Simulation studies, normally, rely on extensive expert knowledge and demanding manual work towards formalizing interactions, establishing models and discovering workflows and processes in those models (Tako and Robinson 2010). The wide-spread use of increasingly more sophisticated sensing devices (Guo et al. 2017; Tricoli et al. 2017; Pilarczyk et al. 2018), and emergence of advanced infrastructures for Information and Communication Technologies (ICT), such as the Internet of Things (IoT) (Gubbi et al. 2013), provide an opportunity for shifting some of the model-building efforts towards the automated derivation of simulation models from data. Could it be that conceptual model-building processes in a typical simulation study need to undergo modifications? Instead of a fully manual determination of a process flow in a model, maybe one needs to only define the key "events" or characteristics, and then build approaches that enable the system to learn how to recognize these key events or characteristics. Such event detection processes can then be followed by an automatic process flow discovery using process mining approaches, such as those presented in (Van Der Aalst 2011).

There are a number of different simulation paradigms, and the most popular being the discrete-event simulation, continuous simulation (also known as system dynamics) and agent-based simulation. Deriving simulation models from data for each of these paradigms can mean different things, and will be followed with different challenges, as we elaborate in this paper. Besides, we also revisit the very standard, and widely adopted and followed steps of a simulation study to reflect these new realities. The paper is structured as follows. In Section 2, we provide background on simulation and overview of some noteworthy efforts in the derivation of simulation models from data. This is followed by analysis of what the data-driven simulation modeling means for each of the simulation paradigms in Section 3. In Section 4, we discuss the implications of these new developments on the advancements in simulation and modeling, and in Section 5, we conclude the paper.

2 BACKGROUND

In the following we provide background on the typical steps involved in a simulation study, to study how they are affected by the new ICT developments and the overwhelming availability of data. Next, we provide an overview of the more significant work that has been done in relation to the data-driven derivation of simulation models.

2.1 Traditional Steps of Simulation Studies

A simulation study, typically, consists of a number of steps, and according to literature, the phase of building a conceptual model precedes the phase of data collection (Maria 1997; Chwif et al. 2013). The conventional steps of a simulation study can be summarized as follows (Banks 2000):

1. Problem formulation (formally defining the problem)
2. Setting of objectives and overall project plan (stating how to approach the problem)
3. Model conceptualization (building a conceptual model, corresponding to the established objectives)
4. Data collection (collecting the data necessary to run the simulation, according to the conceptual model, such as arrival points in time, queuing strategies, service beginnings and completions, etc.)
5. Model implementation (converting the conceptual model into a programming language)
6. Verification (verifying that the conceptual model is implemented correctly)
7. Validation (checking if the implemented model "accurately" represents the real system)
8. Experimental design (designing experiments that target project's objectives, also in terms of accuracy and efficiency)
9. Production runs and analysis (actual running of the simulation and analyzing the output)
10. Document and report (documentation and reporting of results)

Traditionally, a conceptual model is developed based to a large extent on expert knowledge. After a conceptual model is built, decisions on what data needs to be collected are made. With the ease and convenience of collecting data nowadays, which typically precedes the beginning of the simulation study itself, model building and data collection processes need to undergo modifications and redefinitions to make use of the advancements in ICT. The data that is continuously being collected as part of the emergent Internet of Things (IoT) infrastructures can provide significant hints for building conceptual models and, possibly, even automate some parts of the conceptual model building processes. In line with this, in the following subsection, we provide an overview of the advancements and efforts in using data to automatically extract simulation models and process flows. We further claim that these prescribed steps for how to perform a simulation study need to be updated to take advantage and utilize the ongoing data collection processes, which often happen even before a simulation problem is formulated. This implies that steps 3, 4, and 7 (model conceptualization, data collection and validation) need to be revisited, reformulated and/or reordered, as we will elaborate further in the paper, more specifically in Section 4.

2.2 Data-driven Simulation Modeling and Process-Mining

Rozinat et al. in (Rozinat et al. 2009) discuss "discovery of simulation models", where the issue of data-driven discovery of simulation models is brought up, and the notion of "semi-automatic" creation of simulation models is discussed. The authors assume existence of an information system though, which enables some data structuring. The whole process of discovery of simulation models is termed as "process mining".

Process mining is an area that is closely related to the automated derivation of simulation models from data. Namely, process mining is the problem of extracting workflows from event logs. Specialized data mining algorithms are applied to this event log data to identify trends, patterns and details contained in the information system of interest. There has been a substantial body of research that has developed methods for performing process mining, along with tools that have gained a substantial popularity, such as ProM (Van der Aalst et al. 2009). One assumption that process mining approaches make is the availability of event logs that contain time stamps of the relevant beginnings and completions of processes. According to (van der Aalst 2018), any process mining effort starts with a collection of events commonly referred to as an event log, where an event is characterized by:

- A case (also called process instance), e.g., an order number, a patient id, or a business trip,
- An activity, e.g., "evaluate request" or "inform customer",
- A timestamp, e.g., "2015-11-23T06:38:50+00:00", and
- Additional (optional) attributes such as the resource executing the corresponding event, the type of event (e.g. start, complete, schedule, abort), the location of the event, or the costs of an event.

We consider this a significant assumption that can be relaxed using a combination of machine learning approaches to account for datasets that do not follow this event log format, and expert knowledge to identify relevant events (Lazarova-Molnar and Mohamed 2016).

Another similar concept is presented by Hu in (Hu 2011), where the author uses the term "dynamic data-driven simulation". In this case the simulation model is continuously affected by the ongoing real-time data collection. The simulation paradigm that is considered in this work is system dynamics, i.e. continuous simulation, and the application area is wildfire spread. A related discrete-event simulation approach is presented in the PhD dissertation by Akhavian (Akhavian 2015), where the application area is construction processes. The conclusion from this work is that the data-driven simulation models outperform static models created based on engineering assumptions and estimations with regard to compatibility of performance measure outputs to reality. Follow up on this work is presented in (Akhavian and Behzadan 2018). A further encouraging study of the use of data-driven simulation modeling is presented by Kück in (Kück et al. 2016), where the application area is optimization of production scheduling in flexible manufacturing systems.

We see process mining as a significant step in the data-driven derivation of simulation models. We, however, see the data-driven derivation of simulation models as much more than process mining. We see data-driven derivation of simulation models as a methodology that is capable of handling various types of data collections, including time series, where anomaly or event detection processes will need to be performed.

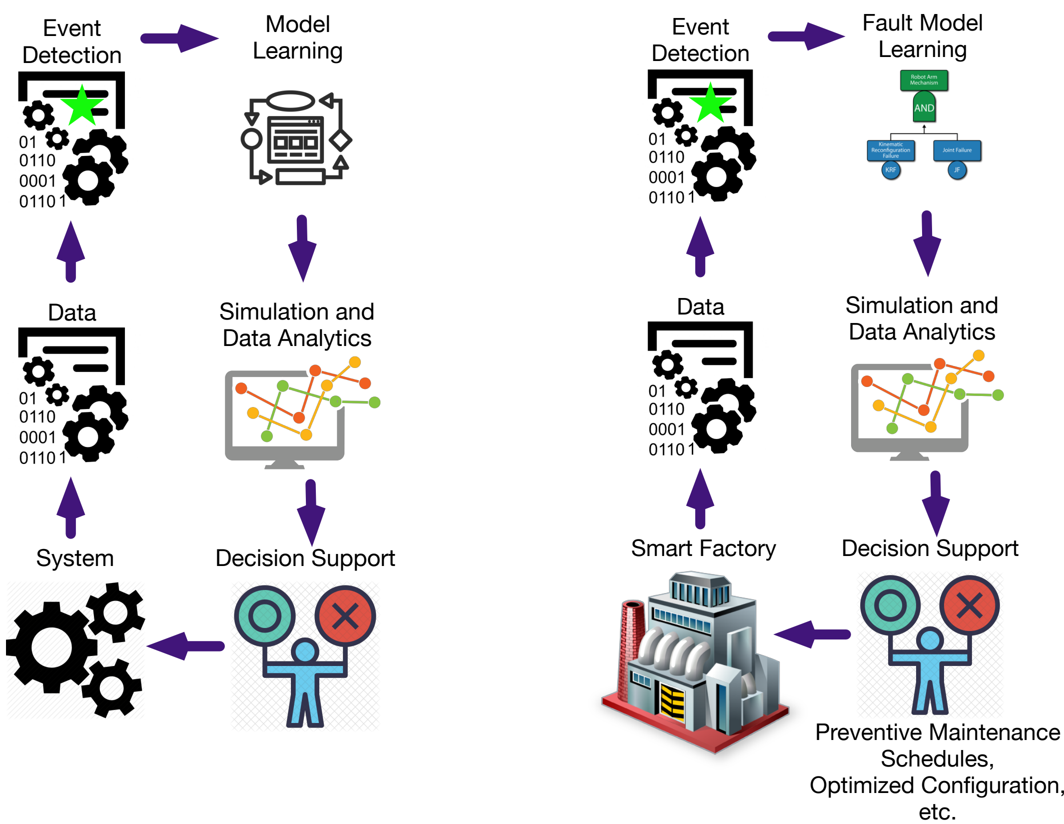
3 DATA-DRIVEN DERIVATION OF SIMULATION MODELS

In the following we describe in detail what data-driven derivation of simulation models can mean for the most popular simulation paradigms. We focus on discrete-event simulation, continuous simulation and agent-based simulation.

3.1 Discrete-Event Simulation

Discrete-event simulation (DES) is the paradigm of choice for systems where system's state changes are abstracted and reduced to only occur at discrete points in time. Typical data collection for DES includes collection of time stamps of events' occurrences. Input data analysis in DES typically includes data collection and distribution fitting (or idealization, in the more general sense) of collected data to determine the structure of the inherent randomness of the relevant events. Typically, input data modeling is preceded by conceptual modeling, a step in which the simulation model is abstracted and the process flow is determined (Robinson 2013). Using process mining approaches and similar, the need for conceptual modeling can be transformed into a need to identify and determine key events and discover ways to describe, characterize and detect them. In some cases, it might be as trivial as a sensors reporting values when a state has changed. However, in many cases, detection of events is not trivial, and more complex event detection methods might be needed to be used (Guralnik and Srivastava 1999), such as methods for anomaly detection (Faigon et al. 2019). Machine learning approaches can be used to this end (Zawbaa et al. 2012; Faigon et al. 2019).

To illustrate the data-driven derivation of discrete-event simulation models, we consider the case of reliability modeling (Lazarova-Molnar et al. 2017; Lazarova-Molnar and Mohamed 2019) alongside a general case, as shown in Figure 1. In the bottom left corner of each of the figures [both (a) and (b)] the system of interest is shown, from where a time series data is collected through sensors and meters. The next step is then the event detection in which relevant events are being detected, which in the reliability case [Figure 1(b)] are faults, failures, repairs' beginnings and completions, etc. (typically extracted using anomaly detection approaches). Then, using process mining approaches or similar, the simulation model of the system is learnt, which is then utilized for decision support for the system of interest through simulation and data analytics. In the reliability case this could mean optimizing maintenance schedules, optimizing components purchase decisions, or optimizing system's configuration. The ongoing data collection allows for continuous validation and calibration of the models, such that they do not need to be explicit and separate steps.



(a) General case (b) Reliability case

Figure 1: Processes involved in data-driven derivation of simulation models for: (a) general case, and (b) reliability analysis.

3.2 Continuous Simulation

Continuous simulation, also often referred to as *system dynamics*, denotes the approach of simulation of systems where state changes are considered to occur continuously in time (Cellier and Kofman 2006). This implies that the state variables that describe the system are continuous. The most adequate way of describing continuous systems is through differential equations, mostly through a set of initial value problems.

In that sense, modeling the functions that describe the rates of change for the various phenomena in a continuous system can be obtained through sensing with high frequency and approximation of the rates of change of measurements by functions. Examples of continuous phenomena would be degradation of materials, or density of pollutants in water. However, if the degradation state of a material or pollutant densities are sensed frequently enough, which is nowadays possible through the highly sophisticated sensing devices, such as nanosensors and similar, the functions of the rate of degradation or the rate of pollution can be easily approximated (Baruah and Dutta 2009; Vikesland 2018). It can, furthermore be continuously adjusted in the simulation model to reflect the changes in the real world.

As mentioned earlier, in Section 2.2, an example approach for the case of data-driven continuous simulation modeling is presented by Hu in (Hu 2011) and illustrated for the case of wildfire spread simulation. The presented simulation system is continually influenced by the real time data streams for better analysis and prediction. Hu identifies the following functions that data-driven simulation enables: dynamic state estimation, online model parameter calibration, and dynamic data-driven event reconstruction. The main disadvantage that the author points out is the complexity of the model, as now it needs not to be oversimplified, thus, it can easily result with a large state space.

As expected, the accuracy of the approximated functions, and the simulation results correspondingly, would depend to high degree on the setup, quality and resolution of the deployed sensors. In that sense, it is of utmost importance that the frequency of sensing and reporting data points is adjusted to capture the relevant state changes, but also not higher than necessary. If machine learning and pattern recognition algorithms are deployed, further relevant and unexpected interdependencies among the different parameters can be extracted as well, which is one of the causes for the increased complexity of derived models, as reported in (Hu 2011).

We illustrate the case of data-driven continuous simulation by population dynamics, which is one of the prominent uses for continuous simulation and it models populations in given geographical regions and their changes over time. Data for populations are usually publicly available, and it is being collected on an ongoing basis. Let us assume that the goal of such simulation study is to decide on some long term strategy pertaining population dynamics. In this case too, the data is already being collected. Therefore, the modeler's task in this case would be to identify variables of interest, and then using methods for data-driven discovery of differential equations (Mangan et al. 2017; Rudy et al. 2017; Schaeffer 2017), identify the differential equations that best describe the dynamics of the population. The advantage of these processes is that once the system's circumstances change and the relationships become affected by these changes, they can be both captured and reflected in the system's simulation. Additionally, some of the assumptions about certain dependencies might become challenged as well, as the new relationships might be triggered by variables that were initially not considered.

3.3 Agent-based Simulation

Agent-based simulation describes the system through an interconnected activity network of agents that are dynamic and have the ability to communicate with each other (Macal and North 2014). Agent-based simulation is utilized for studying crowd behavior, among other uses. The ease of obtaining and installing sensors, combined with their reduced cost and increasing sensing capabilities, enable unlimited possibilities of capturing agents' behaviors and their corresponding relevant states. The popularity of wearable devices and their widespread use present another opportunity for obtaining data needed for various agent-based simulations. In addition, as with the other simulation paradigms, if this data is combined with data analytics and machine learning, new knowledge about behavior interdependencies might be captured and injected into the simulation model. The modeler's effort would then be to identify relevant participants, whose behaviors need to be sensed, as well as incentivize them to participate in the simulation study. One would imagine that privacy ensuring mechanisms will need to be in place for such scenario.

Similar to this, in (Akhavian and Behzadan 2018), the authors report that sensors and an activity recognition methodology based on wearable devices can be used as cost-effective, ubiquitous, and computationally powerful means of enabling data-driven DES models with enhanced reliability over traditional simulation models. Another similar effort has been presented by Keller and Hu in (Keller and Hu 2019), where they present a framework that discovers simulation models in an automated way for mobile agent-based applications, reporting that the results demonstrate that it is possible to discover a variety of interesting models using the framework. Further, in (Porzycki et al. 2015) an agent- based model of pedestrian movement is continuously dynamically built as data is collected. Porzycki et al. also discuss the concept of automatic validation that we see as a significant benefit of the data-driven derivation of simulation models.

As noted previously, one of the typical challenges that agent-based simulation is used for is discovery of emergent behavior (Chan et al. 2010). The big picture of the emergent behavior is obtained through the modeling and simulation of the microscopic agents' behaviors and interactions. To obtain data to automatically build agent-based models, we need a limited number of interacting participants that would agree to be tracked and share their circumstances. Derived models would then be inferred to larger populations using simulation. The fact that a limited number of participating entities are needed implies that the selection of these participants need to be done very carefully, such as to ensure that all types of relevant entities and relevant interactions are covered. Once this has been determined, statistical methods will need to be developed that extract the events and formalize the interactions and interdependencies of agents. Machine learning methods could be utilized to automatically design the different types of agents and their characterizations.

We illustrate the case of data-driven agent-based simulation by a traffic management scenario, as shown in Figure 2. Traffic data is being collected as we are typing, so there is no need to explicitly setup a data collection process. We barely need to observe what data is already being collected, ensure proper privacy mechanisms are in place, and identify and label the main characteristics that are necessary for the objectives of the simulation study, as well as the events and interactions of agents that need to be extracted from the data. Let us assume that the goal is to optimize traffic lights in a part of a city. In that case, the events and characteristics will need to be identified in line with the goal of the simulation study, so one would need to model the driving behavior for different types of vehicles, as well as identify the proportions of the different types of vehicles, which means that agents' (vehicles') technical characteristics will be needed too. As is the case in many countries, Denmark included, most of this information is available for every vehicle by barely entering a registration plate number, but it might be possible to be extracted from the driving behavior as well, by applying machine learning approaches. Therefore, once again, we observe that most of the data is already being collected, and data analytics and machine learning can assist in making it useful for learning simulation models from it. Once all data related decisions have been made, models can be extracted from the data, which can then be utilized for optimizing traffic lights. The advantage of the data-driven model derivation is that models can be automatically calibrated and updated, should circumstances change. Furthermore, models can be also validated on an ongoing basis, thus, enabling more robust simulation solutions than traditional simulation can provide.

4 CHALLENGES AND IMPLICATIONS

Typically, one of the essential and challenging decisions when conducting simulation studies, is to decide on what data, when, and for how long it needs to be collected. With the triviality of obtaining data, these decisions are now transformed into decisions about what datasets to use, and whether the ongoing data collection needs to be supplemented in any way. Furthermore, the input data related decisions now need to be moved to during or after the data collection, or even after preliminary data analytics has been performed. This is so because a lot of data is being collected even before a simulation study is even considered. Through

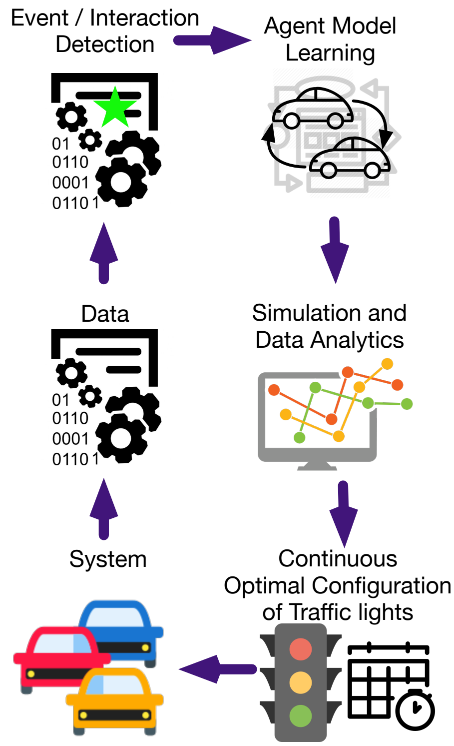
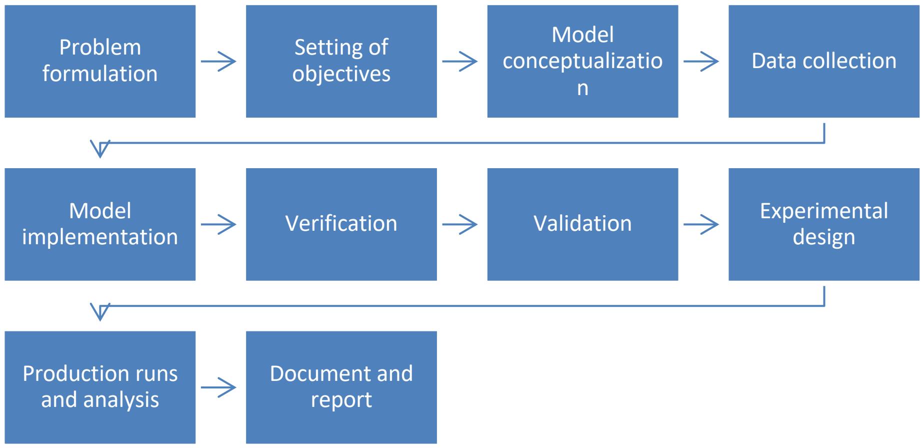


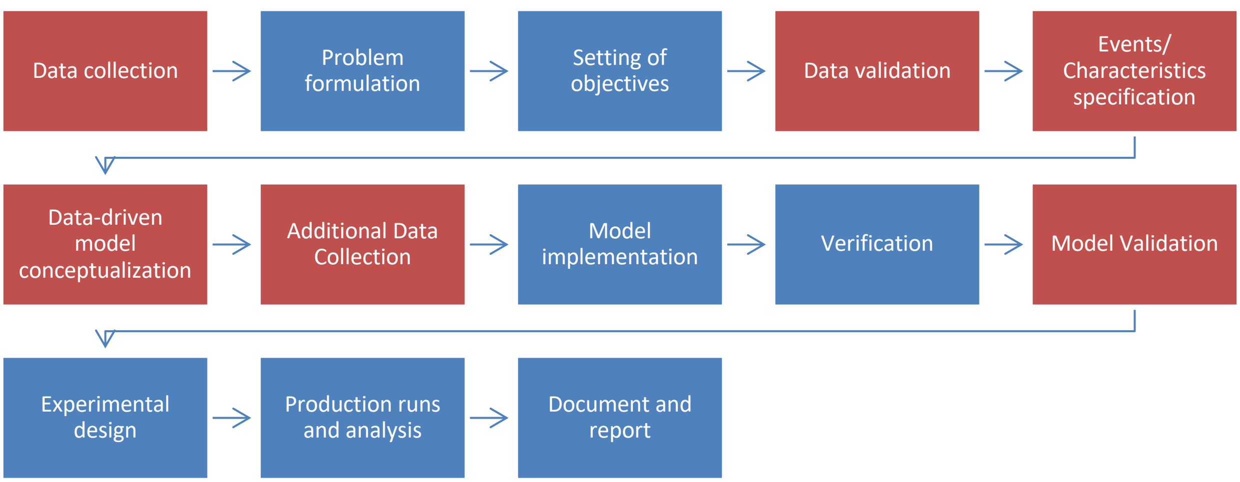
Figure 2: Processes in data-driven agent-based simulation for a traffic management scenario.

delaying these decisions on what data and for how long, there is an opportunity that unexpected behaviors and correlations can already be captured by the ongoing data collection. In this way it can be prevented that over-simplified models are constructed that typically lead to damaging decisions. As expected, the process of automated extraction of simulation model-related data would need to be preceded by a configuration process that will manually determine the processes any key events. Furthermore, the ongoing data collection through various data analytics can itself signal what data is relevant for a given simulation study. However, as pointed out in (Bärring et al. 2018), data acquisition for simulation is a highly complex process, and it is highly related to the existence of standards and meta-models. Therefore, before data-driven derivation of simulation models becomes more widely practiced, more concrete standards and measurements for quality of data need to be in place. As such, a process of data validation will need to be introduced as part of a simulation study. In line with this discussion of the implications of the data-driven derivation of simulation models, in Figure 3, we present both the standard [Figure 3(a)] and revisited steps [Figure 3(b)] for a data-driven simulation study. In the Figure 3(b), data collection processes precede the rest of the steps.

Furthermore, even though validation is presented as a separate step in the predefined order, automatically deriving a simulation model from ongoing data collection provides opportunities to simultaneously perform validation, and, thus, release the need for explicit and separate validation processes. This is especially true if the model is implemented in a modular manner. The key, here, is to ensure that the data is validated, i.e. to ensure that the data correspond to the true measurements and there are no inconsistencies. In Table 1, we identify and summarize the major challenges and benefits related to the data-driven derivation of simulation models for the three most popular simulation paradigms that we have gathered throughout our exploration study. Some of the benefits and challenges are paradigm-specific, whereas some are common, such as the challenge of the increasing complexity that sophisticated models typically carry along.



(a) Steps of a standard simulation study.



(b) Steps of a data-driven simulation study.

Figure 3: Steps of a simulation study: (a) steps of a standard simulation study, and (b) revisited steps for data-driven simulation study (new/updated steps in red).

In addition to the three presented simulation paradigms, there are also simulation approaches that consider hybrid data-driven simulation, such as the one presented in (Saroj et al. 2018), where hybrid traffic simulation model, consisting of a mix of preset and real-time data-driven intersections, is developed. In this case the hybrid is between the data-driven and classic simulation. We are, however, also interested in how data-driven derivation of simulation models can impact other simulation paradigms that are hybrid of the three presented. In line with that, we aim to investigate the impact of data-driven simulation modeling for the proxel-based simulation method as this simulation method builds the state-space of the system of interest on-the-fly (Lazarova-Molnar 2005).

Table 1: Simulation paradigms vs. data-driven simulation modeling.

Simulation Paradigm	Discrete-Event	Continuous	Agent-Based
Challenges	- Event detection algorithms need to be in place - Relevant events need to be manually identified or labeled - Complexity that follows sophisticated models can increase	- High resolution sensors might be necessary for detecting continuous changes - Not all continuous phenomena can be sensed - Complexity that follows sophisticated models can increase	- Mechanisms for ensuring privacy of human participants needed - Mechanisms for quantifying and deriving certain interactions are needed - Complexity that follows sophisticated models can increase
Benefits	- More complex models can become feasible - Validation can be integrated within the ongoing data collection	- Theoretical models can be validated, as well as customized and calibrated to the specific situations - New phenomena can be detected due to ongoing calibration	- Very rich models with high validity can be obtained as agent-based simulation is very complex - Existing theoretical models for agents' interactions can be validated

Finally, we would like to emphasize on the differences between Machine Learning and data-driven Simulation Modeling, as they seem to have a lot in common, especially since Machine Learning is also building (learning) models from data, yet they are very different. Namely, Machine Learning is the prevalent way of learning "black-box" models, albeit, with a substantial shortcoming compared to simulation. One of the significant benefits of simulation, as compared to machine learning is the "white-box-ness", transformed into an *ability to explain*. Namely, when we collect data to build black-box models to represent certain aspects of systems' behaviors, we can build models using less data streams than simulation would need, and, thus, more compactly encapsulate the many intrinsic behaviors and dependencies. This, till now, has been relatively acceptable due to the imitations of the equipment for automated sensing. The limitations of the equipment are, however, steadily changing for the better, and we are now able to obtain much more detailed data that enables a deeper insight into "how" processes are carried out. To illustrate it on the example of smart buildings, some of the latest state-of-the-art smart buildings are equipped with sensors for almost every single piece of equipment. This is the reason why automated deriving of simulation models is becoming an option. Therefore, models with more transparency such as simulation models, in contrast to the black-box models, are becoming an option. The advantage of having simulation models behind the decision support is a greater ability to explain the recommendations, and, therefore, increase their credibility. Explainability and credibility of machine learning-based recommendations has been pointed out as a very critical feature (Doshi-Velez et al. 2017; Zhang and Chen 2018), and simulation might be the answer.

5 CONCLUSIONS

Development of the new ICT, such as IoT and Cloud Computing, enable automated and ongoing collection of relevant data that is directly derived from processes of a system of interest. This availability of data presents a significant game changer in how simulation studies are traditionally performed. In this paper we emphasize the impact that the new advancements in ICT have on the different simulation paradigms, identifying the associated challenges and benefits. The reality is that the data-driven derivation of simulation models is slowly happening and it will definitely impact the ways in which simulation studies are traditionally conducted. The first apparent impacts are in the sequencing of steps of a traditional simulation study, such as the conceptual model building and data collection, which are now drifting away from the original order, as data collection is going to precede all steps, and, if anything, only *additional* data collection might be performed. In such cases, the role of experts will need to be redefend as well, as their role will be more supportive and approving of the data-driven model discovery. Additionally, the ongoing data collection allows for continuous validation and calibration of the simulation models, so they can be performed alongside all steps, instead of being a separate and isolated step. Finally, we expect that the data-driven derivation of simulation models enables development of simulation models that more robust and resilient, and that can withstand and more accurately reflect the changes that occur in the real systems.

REFERENCES

Akhavian, R. 2015. "Data-Driven Simulation Modeling of Construction and Infrastructure Operations Using Process Knowledge Discovery".

Akhavian, R. and A. H. Behzadan. 2018. "Coupling Human Activity Recognition and Wearable Sensors for Data-Driven Construction Simulation". *Journal of Information Technology in Construction* 23(1):1-15.

Banks, J. 2000. "Introduction to Simulation". In *Proceedings of the 2000 Winter Simulation Conference Proceedings*, edited by J. Joines and R. Barton, 9-16. Piscataway, New Jersey: The Institute of Electrical and Electronics Engineers.

Bärring, M., B. Johansson, E. Flores-García, J. Bruch, and M. Wahlström. 2018. "Challenges of Data Acquisition for Simulation Models of Production Systems in Need of Standards". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. Skoogh, N. Mustafee, and A. A. Juan, 691-702. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Baruah, S. and J. Dutta. 2009. "Nanotechnology Applications in Pollution Sensing and Degradation in Agriculture: A Review". *Environmental Chemistry Letters* 7(3):191-204.

Cellier, F. E. and E. Kofman. 2006. *Continuous System Simulation*. Springer Science & Business Media.

Chan, W. K. V., Y.-J. Son, and C. M. Macal. 2010. "Agent-Based Simulation Tutorial-Simulation of Emergent Behavior and Differences between Agent-Based Simulation and Discrete-Event Simulation". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, and J. Montoya-Torres, 135-150. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Chwif, L., J. Banks, J. P. de Moura Filho, and B. Santini. 2013. "A Framework for Specifying a Discrete-Event Simulation Conceptual Model". *Journal of Simulation* 7(1):50-60.

Doshi-Velez, F., M. Kortz, R. Budish, C. Bavitz, S. Gershman, D. O'Brien, S. Schieber, J. Waldo, D. Weinberger, and A. Wood. 2017. "Accountability of AI under the Law: The Role of Explanation". *arXiv preprint arXiv:1711.01134*.

Faigon, A., K. Narayanaswamy, J. Tambuluri, R. Ithal, S. Malmskog, and A. Kulkarni. 2019. *Machine Learning Based Anomaly Detection.* Google Patents.

Gubbi, J., R. Buyya, S. Marusic, and M. Palaniswami. 2013. "Internet of Things (Iot): A Vision, Architectural Elements, and Future Directions". *Future generation computer systems* 29(7):1645-1660.

Guo, S. Z., K. Qiu, F. Meng, S. H. Park, and M. C. McAlpine. 2017. "3D Printed Stretchable Tactile Sensors". *Advanced Materials* 29(27):1701218.

Guralnik, V. and J. Srivastava. 1999. "Event Detection from Time Series Data". In *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, edited by S. Chaudhuri und D. Madigan, 33-42. New York: Association for Computing Machinery.

Hu, X. 2011. "Dynamic Data Driven Simulation". *SCS M&S Magazine* 5:16-22.

Keller, N. and X. Hu. 2019. "Towards Data-Driven Simulation Modeling for Mobile Agent-Based Systems". *ACM Transactions of Modelling and Computer Simulation* 29(1):1-26.

Kück, M., J. Ehm, T. Hildebrandt, M. Freitag, and E. M. Frazzon. 2016. "Potential of Data-Driven Simulation-Based Optimization for Adaptive Scheduling and Control of Dynamic Manufacturing Systems". In *Proceedings of the 2016 Winter Simulation Conference,* edited by P. Frazier, T. Roeder, R. Szechtman, and E. Zhou, 2820-2831. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lazarova-Molnar, S. 2005. "The Proxel-Based Method: Formalisation, Analysis and Applications". PhD Dissertation, University of Magdeburg.

Lazarova-Molnar, S., E. Markoska, and H. R. Shaker. 2017. "Towards Systematic Reliability Modeling of Smart Buildings". In *Proceedings of the 2017 Winter Simulation Conference*, edited by V. Chan, A. D'Ambrogio, G. Zacharewicz, and N. Mustafee, 2484-2494. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Lazarova-Molnar, S. and N. Mohamed. 2016. "Challenges in the Data Collection for Diagnostics of Smart Buildings". In *Proceedings of the Information Science and Applications (ICISA),* edited by K. Kim and N. Joukov, 941-951. Singapore: Springer.

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.

Macal, C. and M. North. 2014. "Introductory Tutorial: Agent-Based Modeling and Simulation". In *Proceedings of the 2014 Winter Simulation Conference*, edited by A. Tolk, L. Yilmaz, S. Y. Diallo, and I. O. Ryzhov, 6-20. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Mangan, N. M., J. N. Kutz, S. L. Brunton, and J. L. Proctor. 2017. "Model Selection for Dynamical Systems Via Sparse Regression and Information Criteria". In *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 473 (2204):20170009.

Maria, A. 1997. "Introduction to Modeling and Simulation". In *Proceedings of the 1997 Winter Simulation Conference*, edited by S. Andradottir, 6-20. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Pilarczyk, K., E. Wlaźlak, D. Przyczyna, A. Blachecki, A. Podborska, V. Anathasiou, Z. Konkoli, and K. Szaciłowski. 2018. "Molecules, Semiconductors, Light and Information: Towards Future Sensing and Computing Paradigms". *Coordination Chemistry Reviews* 365:23-40.

Porzycki, J., R. Lubaś, M. Mycek, and J. Wąs. 2015. "Dynamic Data–Driven Simulation of Pedestrian Movement with Automatic Validation". In *Traffic and Granular Flow'13*, 129-136. Springer.

Robinson, S. 2013. "Conceptual Modeling for Simulation". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, and A. Tolk (PCE), 377-388. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Rozinat, A., R. S. Mans, M. Song, and W. M. van der Aalst. 2009. "Discovering Simulation Models". *Information systems* 34(3):305-327.

Rudy, S. H., S. L. Brunton, J. L. Proctor, and J. N. Kutz. 2017. "Data-Driven Discovery of Partial Differential Equations". *Science Advances* 3 (4):e1602614.

Saroj, A., S. Roy, A. Guin, M. Hunter, and R. Fujimoto. 2018. "Smart City Real-Time Data-Driven Transportation Simulation". In *Proceedings of the 2018 Winter Simulation Conference*, edited by M. Rabe, A. Skoogh, N. Mustafee, and A. A. Juan, 857-868. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Schaeffer, H. 2017. "Learning Partial Differential Equations Via Data Discovery and Sparse Optimization". In *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 473(2197):20160446.

Tako, A. A. and S. Robinson. 2010. "Model Development in Discrete-Event Simulation and System Dynamics: An Empirical Study of Expert Modellers". *European Journal of Operational Research* 207(2):784-794.

Tricoli, A., N. Nasiri, and S. De. 2017. "Wearable and Miniaturized Sensor Technologies for Personalized and Preventive Medicine". *Advanced Functional Materials* 27(15):1605271.

Van der Aalst, W. 2011. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Vol. 2: Springer.

Van der Aalst, W. 2018. "Spreadsheets for Business Process Management: Using Process Mining to Deal with “Events” Rather Than “Numbers”?". *Business Process Management Journal* 24(1):105-127.

Van der Aalst, W. M., B. F. van Dongen, C. W. Günther, A. Rozinat, E. Verbeek, and T. Weijters. 2009. "Prom: The Process Mining Toolkit". *BPM (Demos)* 489(31):2.

Vikesland, P. J. 2018. "Nanosensors for Water Quality Monitoring". *Nature nanotechnology* 13(8):651.

Zawbaa, H. M., N. El-Bendary, A. E. Hassanien, and T.-h. Kim. 2012. "Event Detection Based Approach for Soccer Video Summarization Using Machine Learning". *International Journal of Multimedia and Ubiquitous Engineering* 7(2):63-80.

Zhang, Y. and X. Chen. 2018. "Explainable Recommendation: A Survey and New Perspectives". *arXiv preprint arXiv:1804.11192*.

AUTHOR BIOGRAPHIES

SANJA LAZAROVA-MOLNAR is Associate Professor with the Faculty of Engineering at the University of Southern Denmark. Her current research interests include modeling and simulation of stochastic systems, reliability modeling and analysis, and data analytics for decision support. Since 2019, she is serving on the Board of Directors of The Society for Modeling & Simulation International (SCS). Sanja Lazarova-Molnar obtained her PhD in Computer Science from the University in Magdeburg, Germany, in 2005, specializing in the area of Modeling and Simulation, where she was also a member of the Simulation Group at the Institute of Simulation and Graphics. She has successfully collaborated with industry and academia, authoring more than 80 peer-reviewed research publications. Her email address is slmo@mmmi.sdu.dk.

XUEPING LI is an Associate Professor in the Department of Industrial and Systems Engineering at the University of Tennessee, Knoxville. Dr. Li is the Co-Director of the Health Innovation Technology & Simulation (HITS) Lab (https://hits.utk.edu) and the Director of the Ideation Laboratory (iLab). His research interests include complex systems modeling, simulation and optimization, health information technology and mobile health, data analytics, and smart and connected systems. His research has been sponsored by the National Science Foundation (NSF), National Institutes of Health (NIH), Department of Energy (DOE), Health Resources and Services Administration (HRSA), and so on. He is currently the President of the Modeling & Simulation Division of the Institute of Industrial and Systems Engineering. His email address is xli27@utk.edu. His personal website is http://web.utk.edu/~xli27/.