

INDOLENCE IS FATAL: RESEARCH OPPORTUNITIES IN DESIGNING DIGITAL SHADOWS AND TWINS FOR DECISION SUPPORT

Teresa Marquardt
Catherine Cleophas

Institute of Business
Christian-Albrechts University Kiel
Olshausenstraße 40
24118 Kiel, GERMANY

Lucy Morgan

Lancaster University Management School
Lancaster University
Bailrigg
Lancaster, LA1 4YX, UNITED KINGDOM

ABSTRACT

Digital twins and shadows have gained increasing popularity in industry and research. The terms describe simulation systems that mirror real-world systems, such as service or manufacturing lines, aligned to (near) perfection based on automated data streams. We implement a set of perfect model experiments to demonstrate how deviations between a digital shadow and the real world can arise, affect predictive accuracy, and may be eliminated. As an illustrative example, we simulate a simple sequential production line and its digital shadow. The paper concludes with a summary of identified research opportunities.

1 INTRODUCTION

In recent years, both industry and academia have discovered the idea of digital twins (DT) and digital shadows (DS), combining simulation with big data and IoT devices. Examples, to name but a few, include product design, service design, manufacturing (Tao et al. 2018), urban design (Schrotter and Hürzeler 2020) and even space flight (Shafto et al. 2012). According to Siemens (2021), a closed-loop digital twin aims to keep the simulation and real-world (RW) systems in sync and automatically control the real-world system based on insights from the twin. In that, digital twins constitute what Onggo et al. (2018) term symbiotic simulations. The aim of digital shadows is more modest (Kritzinger et al. 2018): Decision-makers are to predict the effects of interventions from experiments in the highly-valid simulation system and can then decide to implement the most desirable course of action in the real world. The main difference between digital twins and shadows lies in the autonomy of the model: Interventions can be automatic in the twin but require deliberate (human-supported) decisions in the shadow (Rosen et al. 2015). Both concepts share the need for extensive online validation, regardless of whether the supported decisions are automated or not (Onggo et al. 2018).

Indolence can be fatal for both digital twins and shadows: When a relevant value changes in the real-world system but there is no (immediate) update in the simulation, the two systems drift apart. In such cases, automated updates to the real-world systems based on digital twins or decisions supported by digital shadows may be systematically flawed.

As both the underlying ideas and the two concepts' risks are similar, we focus on digital shadows in this paper. Thereby, we neglect how to automatically compute prescriptive insights from a digital twin and focus on the question of automated validity. In an exemplary set of perfect model experiments (Srikrishnan and Keller 2021), we demonstrate the effects of shifts in parameter values and changes in the process structure on a simulation's predictive accuracy. We also explore the success of late updates aimed at re-synchronizing the systems. Specifically, we simulate a sequential production line and its digital shadow. From the outcomes of these experiments and the thoughts outlined in this and the following section, we synthesise a set of research opportunities in the conclusion.

2 BACKGROUND

Seeing the terms digital twin and digital shadow used in numerous ways in the literature, we looked to practitioners for their definition and perceptions. Specifically, we interviewed experts from simulation consultancy Lanner, who provide the predictive digital twin and simulation software WITNESS (Lanner Group Limited 2021). We focus on this set of experts not just for their approachability but also for their overview of a diverse set of industry projects. As a clear limitation, this does not constitute a meaningful empirical survey of expert views on digital twins and shadows. We include conceptualising and implementing such a survey as an opportunity for research in the conclusion of this paper.

The experts point to a recent shift in interest from employing simulation to support strategic decisions to letting it support faster paced tactical or even operational decisions. While not differentiating digital twins and digital shadows, they see the predictive capabilities of simulation models as a major advantage and predict that future enhancements will provide prescriptive capabilities. Over the course of the interview, we identified three areas that sparked interesting research questions:

Granularity causes challenges when combining models that use different levels of detail. When cheap sensors let digital twins synchronize based on real-time data streams as described in Fuller et al. (2020), this raises the question of what level of granularity is appropriate to provide insight.

Agility describes the speed at which the model supports decisions. An agile model can provide fast insights with the most up to date information from the system (Conboy 2009). By design, digital twins and shadows pull live data from the real-world system into the digital model and can thus provide greater agility than traditional simulation systems (Aheleroff et al. 2021). The specific function of the model raises the question of how to combine live and historic data to provide agile insight. For example, a predictive model will require some combination of live and historic data, which raises the question of which data to use and when to gain the most relevant insights. Due to the automated nature of controls, digital twins are more agile, but also more vulnerable to systematic faulty decision making when validity is not ensured.

Validity describes both the impact of correctly identifying and analysing input parameter values and the impact of making a poor model assumption (Fagiolo et al. 2019). Both of these fall under validation of the inputs in the framework proposed by Nelson (2013). We do not consider validation of the model logic, such as system layout, within this paper. Of course the alignment of the twin and the real-world is highly dependent on the purpose of the model (Troitzsch 2009). Input validation has been considered as a challenge of simulations in general in the past (see Sargent (2020) and Law (2019)) and emerged from our discussions with experts as an even more crucial aspect when considering digital twins. In the experts' opinion, this problem is worse for a model aiming to making operational decisions.

Delving deeper into the literature on validity in conjunction with digital twins and shadows, there are a vast number of offline validation techniques. Traditional methods aim to compare the simulation's output measures to what occurred in the real-world. For a summary of general methods see Sargent (2020). More generally, Kleijnen et al. (1998) propose a regression-based test on the differences between the simulated and real-responses for validation of a trace-driven simulation. Khan et al. (2018) propose an offline validation approach for digital twins, based on model based testing for legacy systems.

Perhaps more relevant to digital twins is validating the behaviour of the simulation model online, through time. Lugaresi et al. (2019) propose a signal processing approach for the validation of online (real-time) simulation models for manufacturing systems. They introduce the quasi-trace-driven simulation approach. Morgan and Barton (2021) also consider simulation trace information looking to discriminate between number in system trajectories with distinct utilisation or dynamic behaviour by introducing a Fourier based statistic.

3 EXEMPLARY PERFECT MODEL EXPERIMENT

We aim to showcase the effects when a change in the real-world system causes it to deviate from the digital shadow. To that end, we create a so-called perfect model experiment as described in Srikrishnan and Keller

(2021), where both the real world and its digital shadow are actually simulations and deviations can be manipulated at will. We implement the model in Python 3.9, using the package SimPy to implement it in the discrete-event-based paradigm (Team SimPy 2021).

3.1 Model Description

The system represents a production process with two sequential manufacturing steps and two interspersed quality control steps. Each step relies on one or more specific resources, termed R1, R2, R3a, R3b, and R4, representing machines or manual work. Figure 1 illustrates the process as a flow chart, highlighting potential queues (rectangles) in front of resources (circles) as well as decision splits and merges (triangles).

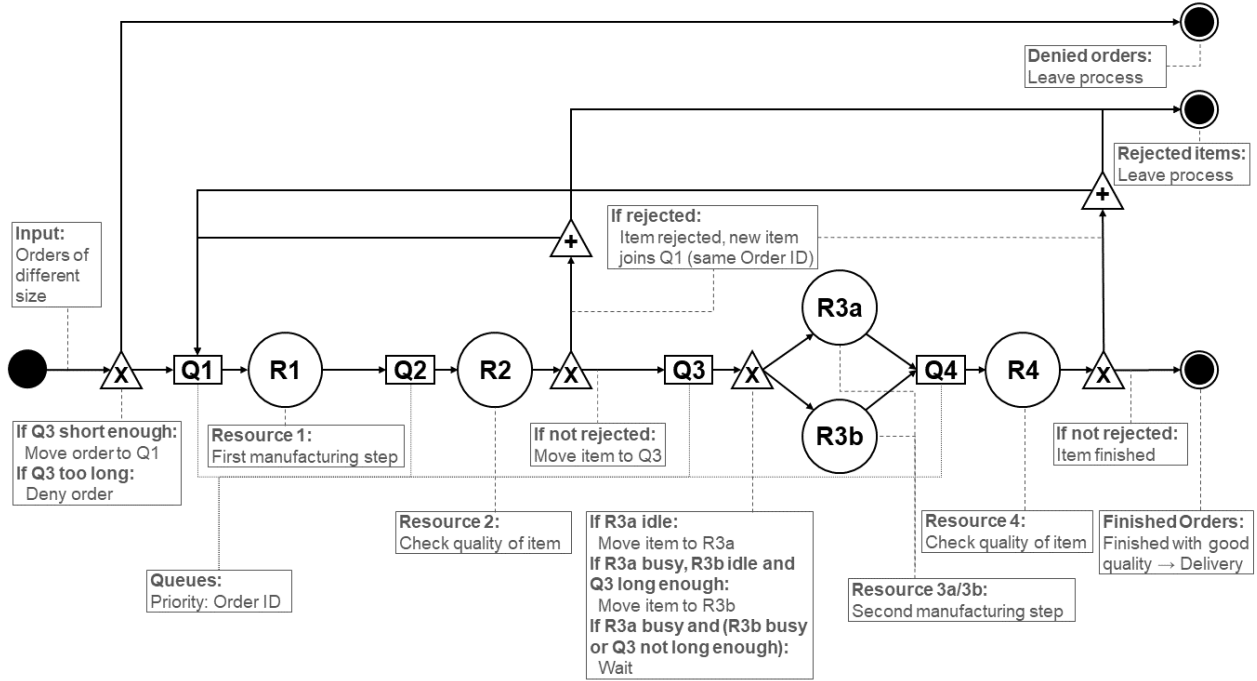


Figure 1: Simulation Model.

The process is driven by orders that include varying numbers of items. On arrival, all items from a given order enter the first queue ahead of the first manufacturing step as implemented by R1. All resources can only process one item at a time. All queues are processed according to a priority discipline based on the initial arrival of the order to the system. The same logic applies as items queue up for quality control at the resource R2. If the length of the third queue indicates that the system is overloaded, the system rejects newly arriving orders. Orders are accepted again when the queue length falls below a given threshold.

Both manufacturing steps are subject to a given error rate and flawed items do not pass quality control. If a quality control step rejects an item, the item needs to be re-produced and thus joins the first queue again. However, as each item is prioritised according to the arrival time of its related order, it can then overtake later items already waiting in the queue.

Items that do pass the first quality control progress to the second manufacturing step. Note that this step can involve two resources, R3a and R3b. A scheduler assigns items primarily to R3a; items are passed to R3b if the queue length reaches a given threshold in terms of items, R3a is busy, and R3b is available. Resource R3a and R3b can differ in their processing time, the number of parallel items that they can handle, and error rates. Another quality control step follows, utilizing resource 4. When all items belonging to an order have been manufactured to an acceptable quality standard, the order leaves the system.

Table 1: Simulation parameters for the baseline scenario.

Parameter	Distribution	Base Value
Order arrival	exponential distribution	$Exp(1/33)$
Items per order	exponential distribution	$Exp(0.25)$
<i>Manufacturing Step 1</i>		
processing time R1	exponential distribution	$Exp(0.5)$
error rate R1	uniform	0.1
<i>Quality Control Step 1</i>		
processing time R2	deterministic	1
<i>Manufacturing Step 2</i>		
processing time R3a	exponential distribution	$Exp(1/6)$
error rate R3a	uniform	0.1
Q3 R3b activation threshold	deterministic	8
availability R3b	deterministic	TRUE
processing time R3b	exponential distribution	$Exp(1/8.5)$
error rate R3b	uniform	0.2
Q3 system closure threshold	deterministic	15
<i>Quality Control Step 2</i>		
processing time R4	deterministic	1

The resulting simulation model features sets of parameters that govern all steps. Table 1 lists the simulation parameters and the corresponding settings for the base scenario. Our choice of parameters in the base scenario reflect a system where there is one preferred resource (R3a) with higher processing speed and quality. We set the split flow threshold at R3a to eight as, under this policy, we see stable throughput times for items flowing through the system. Resource R3b is utilised regularly ($\approx 25\%$ of the time) and there are still times when orders are rejected due to short-term system overload.

When digital shadow (DS) and real-world system (RW) are aligned, both work according to the same parameter values and differ only in values drawn from random distributions, such as the order arrival times. In this scenario, results from DS and RW only differ in stochastic variation. Optionally, modellers can even exclude such random variation by setting the same random seed for both sets of simulation experiments. In the real world, this would be equivalent to running the digital shadow with data from a sensor facilitated live data feed.

3.2 Scenarios and Experiments

Experiments simulate the manufacturing process over the course of 30,500 time units, assuming that the system is empty at the start of the time horizon. To account for stochastic variability, we run the digital shadow simulation over 100 replications. We split the time horizon into a burn-in period of 500 time units, and three distinct sequential stages of 10,000 time units each (termed I, II, III in the following) for most experiments. For an experiment on faulty updates, we add a Stage IV, hence dividing the time horizon after the burn-in period into four stages of 7,500 time units. Stage I represents the stable system state after the burn-in period. To cause the real-world system to deviate from the digital shadow, we change parameter values or distributions in the respective RW setting in stages II to IV. To align or re-align the two systems, we adjust the same value in the digital shadow, either in the same or in a later stage. Hence, the real world may change in Stage II, with an adjustment to the digital shadow following in step III, and another adjustment to DS in Stage IV. Notably, Stage IV and the related part of the time horizon are not included in each experiment, as three stages suffice to model simpler updating sequences.

Changing parameters even in this simple model can generate a combinatorially large number of experiments. To examine the deviations due to invalid assumptions and shocks, we showcase changes to resource R3b – deviations in parameters or distributions pertaining to any other system entity would be just as feasible. Here, we pick R3b as it is set-up to be the system’s bottleneck, e.g., when determining whether new orders are accepted or not. For each deviation, the digital shadow is either perfectly updated, updated late, incorrectly updated, or not updated at all. The following paragraphs describe parameters we control with respect to R3b and detail on how we change them within our experimental scenarios.

Resource Availability: A Boolean parameter determines the general availability of the resource R3b. In the base scenario, R3b is available and can be activated when needed. Note that even when both resources work at the same speed, making R3b available differs from increasing the capacity of R3a, as R3b is only activated when the queue length reaches a given threshold in terms of items, as motivated, for example, by cost considerations.

Resource Processing Speed: When deviating on speed, we reduce the processing speed of resource R3b to 0.6 times the base scenario speed. This might represent replacing a worker with a less skilled substitute at short notice.

Resource Processing Distribution: We also test the implications of deviating assumptions in the digital shadow by letting the processing speed of resource R3b follow a different distribution in DS versus RW. For example, if R3b relies on human workers, two alternating staffers with diverging speeds might work in rotation, creating a bi-modal distribution. We represent this by drawing processing speeds from two overlapping beta distributions in RW, while letting DS follow the assumption of the exponential distribution of speeds that applies in the base scenario. Specifically, we draw the processing speeds in the RW setting from $X \sim \text{Beta}(2,6)$ (expected value is 0.25) and mirror every second value at 0.5. The values are then linearly transformed to match the mean of the exponential distribution used in DS.

Resource Activation Threshold: We vary the scheduling behaviour such that, in RW, resource R3b is not activated for a queue length of 8 items, as in the base scenario, but only for a queue length of at least 11 items. This could constitute an update in the scheduling logic of the real world that the digital shadow does not, initially, implement.

3.3 Indicators of System State

To feedback the system state, we compute the throughput time per last item while also monitoring queue lengths. As both indicators follow a similar pattern over time in the showcased experiments, the analysis presented here focuses on the length of queue 3 for parsimony.

For all indicators, we compute simulation means per time in the simulation by averaging the observed replication values across all 100 runs. The mean across replications and the individual output replications were used to compare the impact of deviations between the real world and the digital shadow.

4 SHOW CASE EXPERIMENTS

Validating the Base Scenario: A valid digital shadow accurately models the real-world system’s structural characteristics, distributions, and parameter values. To validate the perfect model experiment, we simulate the digital shadow and the real-world settings with the same parameter values and set the pseudo-random-number generator to the same seed. In consequence, performance indicators from both settings are absolutely congruent (not shown). In practice, digital shadows are built to provide insight to business challenges. A perfect fidelity model comes with additional complexity and often a simpler model can provide better user insight (see Tako et al. (2020) and references within). The digital shadow could therefore represent a simplified model of the real world and deviations should be expected between the performance indicator outputs even when data streams are perfectly synchronized.

Even when the digital shadow is a perfect fidelity model, as in our experiments, using it for predictive purposes relies on calibrated stochastic distributions, and thus a perfect match should not be expected. We examine this by setting different seeds for the RW and DS base scenarios while keeping distribution parameters aligned. This yields a slight, random divergence in performance indicators around 1% (compare Figure 2). newAs expected, a test for the difference in the mean length of queue 3 across all stages, finds no significant difference in performance between DS and RW (p-value 0.342).

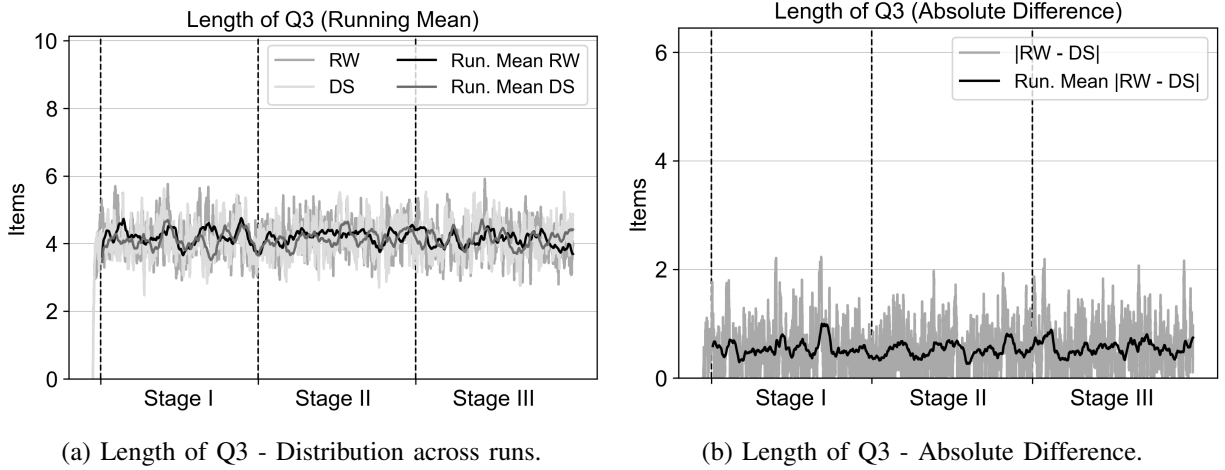


Figure 2: Validating the Base Scenario – Real World and Digital Shadow are in Alignment Across Three Stages.

Structural Shocks: To test the effects of a structural shock, we change the availability of resource R3b in the RW setting. This set of experiments represents, for example, a case where a worker leaves their station in the real world. If the digital shadow is updated perfectly and instantly, the effect in terms of reduced manufacturing capacity is visible in indicators from both settings (not shown). However, if the availability of R3b is not monitored, there is no corresponding update in the digital shadow and the systems deviate. For example, if R3b was a human worker, this type of deviation may arise when regulations prohibit the firm to monitor when workers take planned or unplanned breaks within their shifts. Figure 3 illustrates the resulting deviation when the resource becomes unavailable in the real world in stage II and returns in stage III without a corresponding update in the DS. In consequence, the absolute difference of the mean length of Q3 increases to up to 4.92 items. A test for the difference in the mean length of queue 3 across all stages finds a significant difference in performance between the digital shadow and real-world system as expected (p-value 8.95×10^{-39}). Notably, as the resource becomes available again, it appears that, in this system, it takes only few time units for the divergence to return to its regular volume as caused by random variation. A test for the difference in the mean length of queue 3 in stage III finds no significant difference in performance between DS and RW (p-value 0.947). Repeating the same test with a burn-in of 1,000 time units in stage III gave the same conclusion with an even higher p-value (p-value 0.994). Figure 4 shows that the divergence can also be regulated by updating DS, even when the update is implemented late. In the corresponding experiment, the resource becomes unavailable in RW in stage II and is set unavailable in DS in stage III. The increased deviation seen in stage III of Figure 4 is a symptom of increased variation due to higher traffic intensity when the resource is unavailable. As expected, a test for the difference in the mean length of queue 3 across all stages finds a significant difference in performance between DS and RW (p-value 1.09×10^{-32}). However, the average relative difference in the length of queue 3 across systems returns to $< 2\%$ in stage III (which is similar to the level seen in stage I) after exceeding 65%

in stage II, when the deviation was known to occur. Additionally, in this experiment, the system quickly settles back into steady state after the correction. A test for the difference in the mean length of queue 3 in stage III alone, conducted using 100 replications of each setting, concluded no significant difference in performance between DS and RW (p-value 0.395). Repeating the same test with a burn-in of 1,000 time units in stage III leads to the same result with an even higher p-value (p-value 0.467).

Note that the quick return to steady state seen in this experiment is not guaranteed for all digital shadows. This prompts the question of whether additional burn-in periods are required after a deviation has been identified and then corrected for, and if so, how long a burn-in period needs to be.

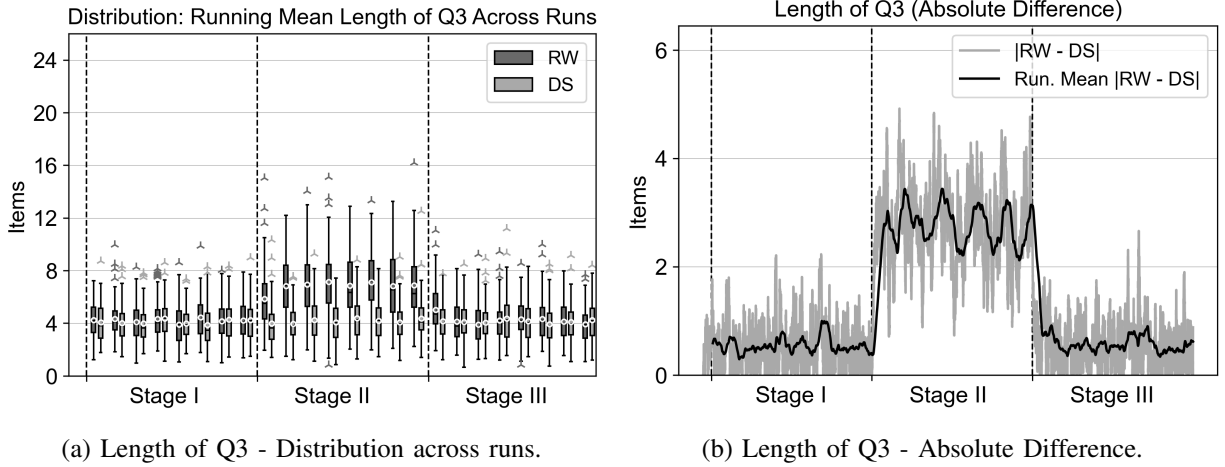


Figure 3: Worker leaves in Stage II and returns in Stage III - No update of the digital shadow.

Diverging Assumptions: In another set of experiments, we gauge the effect of diverging assumptions with regard to the random distributions underlying the speed of resource R3b. To that end, we let the RW processing speed follow a bi-modal distribution, whereas DS implements an exponential distribution calibrated to have the same average speed. In this set of experiments, the real world and the digital shadow rely on the same seed. Notably, the deviation between the performance indicators from the real world and the digital shadow is not significantly larger when the distributions are not aligned (not shown). A test for the difference in the mean length of queue 3 across all stages finds no significant difference in performance (p-value 0.051). We conclude that such differences between RW and DS can be difficult to detect using aggregate indicators and may lead to fallacies for experiments aiming at predicting the outcomes from system changes. A wider investigation of the impact of diverging or incorrect assumptions is required. This is of particular interest in systems with frequent changes where data to model the input distributions is scarce. Related, Zhou and Liu (2018) consider uncertainty quantification for simulation models in an online context.

Parameter Value Shocks: To test the effects of a shock in terms of a deviating parameter value, we reduce the speed of resource R3b by 40% in the RW setting. This set of experiments represents, for example, a case where a less experienced substitute worker takes over the station in the real world. Once more, updating the digital shadow in time causes the effect to manifest in indicators from both settings (not shown). The change in performance indicators shows that a change in the parameter value impacts outcomes less than a structural change of the related resource. Thus, the average length of Q3 increases by only 0.7 units on average. Figure 5 shows the implications when DS is initially not updated: The increase in queue length observed in the real world is not mirrored by DS in Stage II, causing a corresponding absolute difference between the performance indicators from both settings. A test for the difference in the

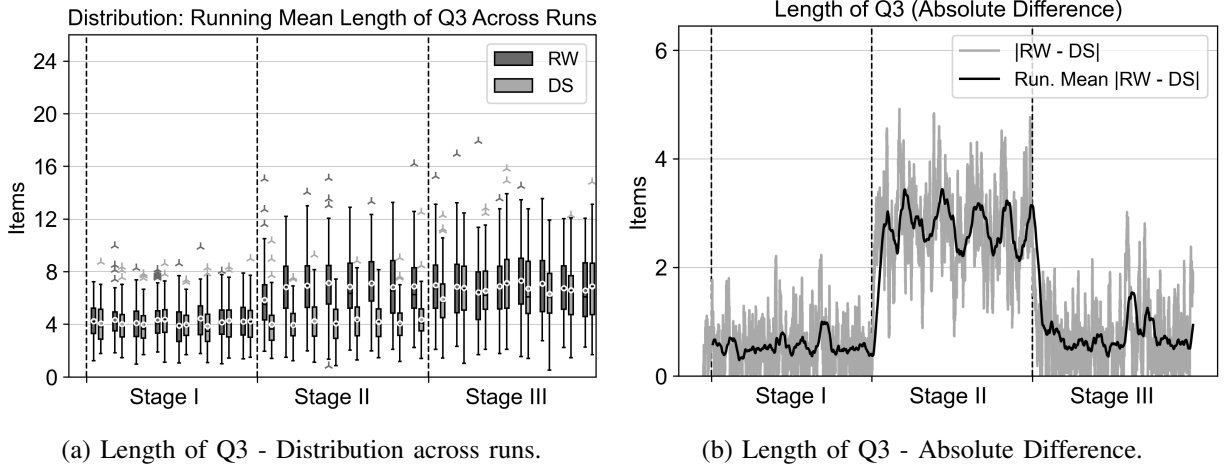


Figure 4: Worker leaves in Stage II - Late update of the digital shadow in Stage III.

mean length of queue 3 across all stages finds a significant difference in performance between DS and RW ($p\text{-value } 6.27 \times 10^{-7}$). When updating the digital shadow in Stage III, the divergence quickly returns to its baseline. A test for the difference in the mean length of queue 3 in stage III alone finds no significant difference in performance ($p\text{-value } 0.631$).

Faulty Digital Shadow Updates: Finally, we offer a set of experiments to exemplify how faulty updates can only superficially re-align the digital shadow. To that end, we add stage IV, such that stages II to IV each include 7,500 time units. The experiment underlying Figure 6 traces the development of performance indicators when, in the real world, resource R3b slows down in stage II, but the update of the digital shadow in stage III is based (incorrectly) on the assumption that R3b is only activated for a longer queue length. In stage IV, the incorrect assumption is corrected for and the digital shadow is re-aligned with the real world. As expected with regard to parameter value shocks, the change in processing speed causes a relatively small but significant deviation. A test for difference in the mean length of queue 3 within stage II conducted using data from 100 replications of each system concludes that there is a significant difference in mean queue length ($p\text{-value } 3.17 \times 10^{-11}$). Changing the queue length required to activate R3b appears to ameliorate these differences, however, the analogous test for a difference in means looking at stage III also finds a significant difference ($p\text{-value } 0.0172$). When only considering performance indicators visually, modellers might conclude that the update was appropriate. There is no evidence of a significant difference in the length of queue 3 during stage IV after correcting the update by re-aligning the required queue length and the speed of R3b ($p\text{-value } 0.880$).

5 DISCUSSION AND CONCLUSION

To enable a simulation to serve as a digital shadow, even more so to serve as a digital twin, modellers have to ensure that it accurately represents the state of the corresponding real-world system. After highlighting some challenges of digital twin design, this paper used a perfect model experiment to illustrate the potential consequences of *indolence*, when the digital shadow lags behind changes in the real-world system. Specifically, we pinpointed ways in which the digital shadow can deviate from the real-world system due to shocks or faulty assumptions.

Shocks in Parameter Values or Structure: Any shock in the real-world system has to be immediately mirrored by the digital shadow to keep the two aligned. As our experiments show, when parameter updates are synchronous, changes in the magnitude of deviations between real world and digital shadow only occur

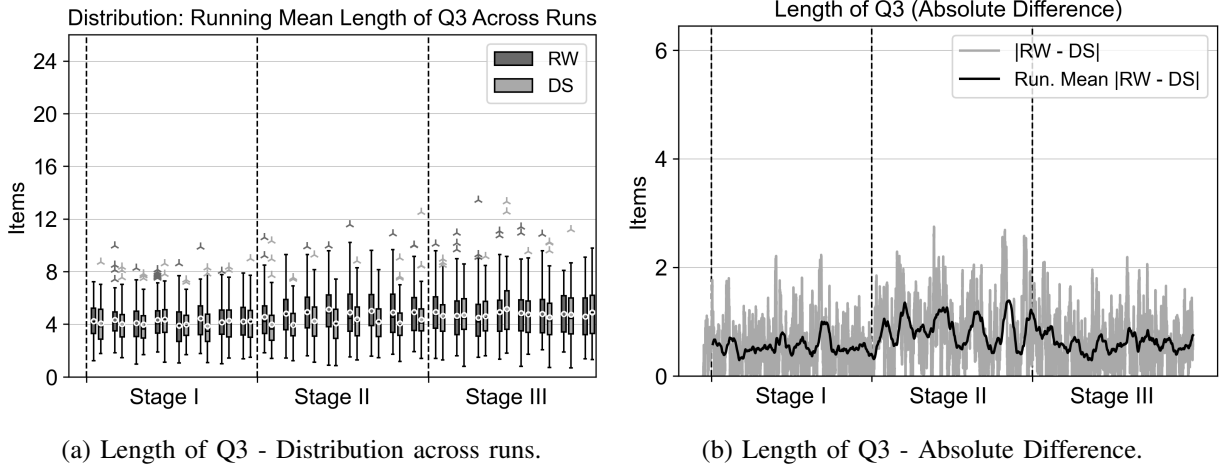


Figure 5: Worker slows down in Stage II, late update of the digital shadow in Stage III.

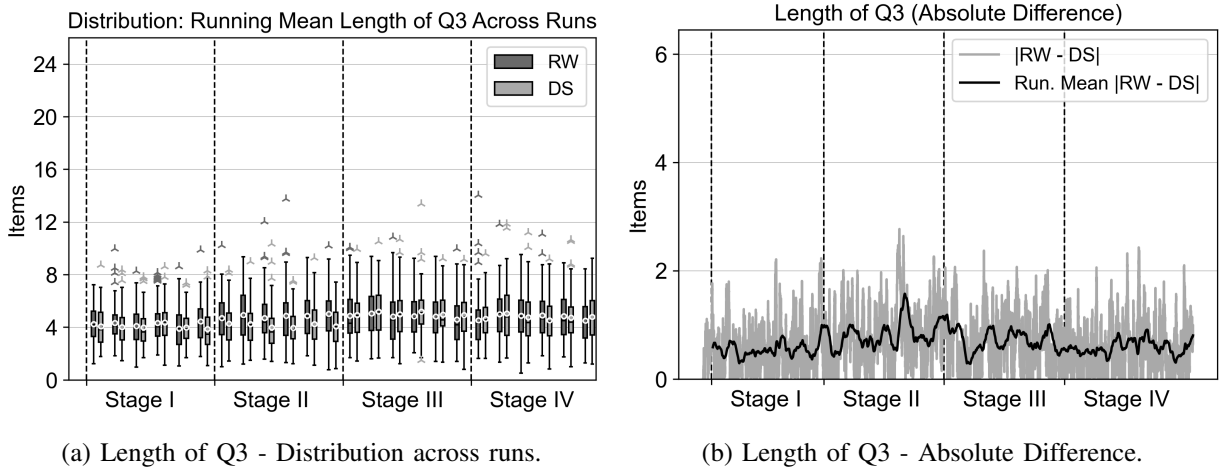


Figure 6: Worker slows down in Stage II, faulty update of activation threshold in Stage III, correct update in Stage IV.

when the shock also affects the level of random variation. However, depending on the flow of information, the digital shadow may not immediately update when the real-world system changes. In this case, the two systems only align if the change in the real world is undone, e.g. when the resource becomes available again or when the digital shadow is belatedly updated. In either case, the shift over time requires another return to steady state, similar to the initial burn-in period of the digital shadow analyzed here.

Diverging Assumptions: Our showcase experiments illustrated that comparing the real world and the digital shadow indicators superficially and in a steady state may not suffice to identify deviations that result from diverging assumptions. For example, when only the shape of a distribution, but not its mean differ, deviations are not obvious in aggregate indicators. This could have negative effects when using a digital twin to guide control decisions for the real-world system. Furthermore, it is not always immediately clear what has truly changed in the real world. When erroneously changing a resource activation rule rather than its processing speed, the systems appeared to be re-aligned in the experiment presented in Figure 6.

Our showcase was limited to investigating a simplified combination of real-world system and digital shadow. Potential further model structures that we could not explore due to the limited scope of this paper include more activities, a more varied order of resources, and variant products with different manufacturing steps. Beyond the discrete-event-based paradigm, systems with a human component may be more accurately represented by an agent-based model. These are, however, notoriously harder to validate, as pointed out in Windrum et al. (2007).

There is plenty of potential for future work in the area of implementation and use of digital shadows and twins. Potential directions include online validation techniques for identifying when the two systems have diverged. Furthermore, procedures for quantifying and correcting for bias in performance indicators after a correction to a digital twin has been made (analogous to burn-in bias) are needed. Additionally, techniques for uncertainty analysis may be required to understand uncertainties emerging from incorrect distributional assumptions in the digital shadow and potential changes in distribution through time from the real world. Furthermore, information systems engineering needs to consider the rate at which to pull data from the real world and update the digital shadow. This was not a consideration of our example but would be if data were pulled every x time units from the real world. The above points are traditional simulation problems, that have to be viewed in a new light given the challenges presented in implementing a digital twin. With this challenge comes the opportunity to use techniques from machine learning and big data to provide automated insights. Last but not least, while the research presented here was inspired by interviews with simulation experts from a single firm, the rising importance of digital twins in the industry also calls for a more deliberate empirical study. Such an endeavour could systematically investigate the assumptions and expectations that relate to the terms digital twin and digital shadow in the industry. In that, researchers should differentiate firms interested or engaged in applying the concept to model their value creation process and firms that create value by developing simulation systems and offering related consulting services.

ACKNOWLEDGEMENTS

We thank simulation experts from Lanner for the discussion and interview that motivated this investigation.

REFERENCES

- Aheleroff, S., X. Xu, R. Y. Zhong, and Y. Lu. 2021. "Digital twin as a service (DTaaS) in industry 4.0: an architecture reference model". *Advanced Engineering Informatics* 47(2):101225.
- Conboy, K. 2009. "Agility from first principles: Reconstructing the concept of agility in information systems development". *Information systems research* 20(3):329–354.
- Fagiolo, G., M. Guerini, F. Lamperti, A. Moneta, and A. Roventini. 2019. "Validation of agent-based models in economics and finance". In *Computer simulation validation*, edited by C. Beisbart and N. J. Saam, 763–787. Cham: Springer.
- Fuller, A., Z. Fan, C. Day, and C. Barlow. 2020. "Digital twin: Enabling technologies, challenges and open research". *IEEE Access* 8:108952–108971.
- 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 (CASE)*, edited by F.-T. Cheng, M. Dotoli, J. Li, J. Li, D. Popa, K. Saitou, C. Seatzu, B. Vogel-Heuser, X. Xie, and Q. Zhao, 421–426. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Kleijnen, J. P., B. Bettonvil, and W. Van Groenendaal. 1998. "Validation of trace-driven simulation models: a novel regression test". *Management Science* 44(6):812–819.
- Kritzinger, W., M. Karner, G. Traar, J. Henjes, and W. Sihn. 2018. "Digital Twin in manufacturing: A categorical literature review and classification". *IFAC-PapersOnLine* 51(11):1016–1022.
- Law, A. M. 2019. "How to build valid and credible simulation models". In *2019 Winter Simulation Conference (WSC)*, edited by N. Mustafee, K.-H. G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, 1402–1414. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Lanner Group Limited 2021. "Lanner. Future. Proof". <https://www.lanner.com/>, accessed 9th April.

- 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 (CASE)*, edited by F.-T. Cheng, M. Dotoli, J. Li, J. Li, D. Popa, K. Saitou, C. Seatzu, B. Vogel-Heuser, X. Xie, and Q. Zhao, 450–455. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Morgan, L. E., and R. R. Barton. 2021. "Fourier Trajectory Analysis For System Discrimination". *European Journal of Operational Research*.
- Nelson, B. 2013. *Foundations and methods of stochastic simulation: a first course*. New York: Springer Science & Business Media.
- Onggo, B. S., N. Mustafee, A. Smart, A. A. Juan, and O. Molloy. 2018. "Symbiotic simulation system: Hybrid systems model meets big data analytics". In *2018 Winter Simulation Conference (WSC)*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1358–1369. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Rosen, R., G. Von Wichert, G. Lo, and K. D. Bettenhausen. 2015. "About the importance of autonomy and digital twins for the future of manufacturing". *IFAC-PapersOnLine* 48(3):567–572.
- Sargent, R. G. 2020. "Verification And Validation Of Simulation Models: An Advanced Tutorial". In *2020 Winter Simulation Conference (WSC)*, edited by K.-H. G. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, 16–29. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Schrotter, G., and C. Hürzeler. 2020. "The digital twin of the city of Zurich for urban planning". *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 88(1):99–112.
- Shafit, M., M. Conroy, R. Doyle, E. Glaessgen, C. Kemp, J. LeMoigne, and L. Wang. 2012. "Modeling, simulation, information technology & processing roadmap". *National Aeronautics and Space Administration* 32:1–38.
- Siemens 2021. "Digital Twin". <https://www.plm.automation.siemens.com/global/en/our-story/glossary/digital-twin/24465>, accessed 9th April.
- Srikrishnan, V., and K. Keller. 2021. "Small increases in agent-based model complexity can result in large increases in required calibration data". *Environmental Modelling & Software* 138:104978.
- Tako, A. A., N. Tsiptsias, and S. Robinson. 2020. "Can we learn from simplified simulation models? An experimental study on user learning". *Journal of Simulation* 14(2):130–144.
- Tao, F., J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui. 2018. "Digital twin-driven product design, manufacturing and service with big data". *The International Journal of Advanced Manufacturing Technology* 94(9):3563–3576.
- Team SimPy 2021. "SimPy Discrete event simulation for Python". <https://simpy.readthedocs.io/en/latest/>, accessed 9th April.
- Troitzsch, K. G. 2009. "Not all explanations predict satisfactorily, and not all good predictions explain". *Journal of Artificial Societies and Social Simulation* 12(1):10.
- Windrum, P., G. Fagiolo, and A. Moneta. 2007. "Empirical validation of agent-based models: Alternatives and prospects". *Journal of Artificial Societies and Social Simulation* 10(2):8.
- Zhou, E., and T. Liu. 2018. "Online quantification of input uncertainty for parametric models". In *2018 Winter Simulation Conference (WSC)*, edited by M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, 1587–1598. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

AUTHOR BIOGRAPHIES

TERESA MARQUARDT studies Business Administration and is a Student Research Assistant of Catherine Cleophas at CAU Kiel University. Her research interests are statistical data analysis, visualisation, and simulation modelling. Her email address is teresamq.tm@gmail.com.

LUCY MORGAN is a Lecturer in Simulation and Stochastic Modelling in the Department of Management Science at Lancaster University. Her research interests are input uncertainty in simulation models and arrival process modelling. Her e-mail address is l.e.morgan@lancaster.ac.uk.

CATHERINE CLEOPHAS is Professor of Service Analytics in the Department of Business Administration at CAU Kiel University. Her research interests are decision support through predictive models in general and simulation in particular. Her email address is cleophas@bwl.uni-kiel.de.