The International Journal of Advanced Manufacturing Technology (2023) 124:2423–2435
https://doi.org/10.1007/s00170-022-10674-6

ORIGINAL ARTICLE

Simulation-based digital twins monitoring: an approach focused on models' accreditation

Carlos Henrique dos Santos[1] · Afonso Teberga Campos[1] · José Arnaldo Barra Montevechi[1] · Rafael de Carvalho Miranda[1] · João Victor Soares do Amaral[1] · José Antonio de Queiroz[1]

Received: 15 September 2022 / Accepted: 6 December 2022 / Published online: 13 December 2022

Abstract
This work proposes the development of a tool focused on monitoring simulation-based digital twins (DTs). Although it has been a popular approach in recent years, adopting DT simulation models for decision support in production systems implies some challenges, emphasizing the need for techniques to ensure the accreditation of the models over time. In this case, since the literature does not present alternatives for evaluating DTs models during their operation, this work proposes a tool based on the K-near neighbors (K-NN) classifier and the *p* control chart. In this way, it is possible to compare the behavior of the physical and virtual environments, monitoring the DT to guarantee its accreditation to support decisions. To assess the applicability of the proposed approach, we applied it to two real study objects with different characteristics, where the tool was able to evaluate the DTs simulation models. As a result, we observed that the proposed tool acted as a supplement to the DTs in order to make them more robust and reliable. Moreover, it was possible to monitor DTs with several evaluation variables in a simple way and with a friendly interface.

Keywords Digital twin · Simulation · Accreditation · Monitoring

1 Introduction

The Industry 4.0 era has driven several developments aimed at more efficient decision-making in manufacturing processes [1]. In this case, in addition to the evolution of information technologies (IT), several solutions has been developed for increase the systems efficiency, such as the so-called Internet of Things (IoT), Big Data, and Cloud Technology, which allow information sharing between smart devices, complex data analysis, and data storage and processing without physical resources, respectively [2]. Furthermore, Zhong et al. [3] highlight the so-called cyber-physical systems (CPS) as a fourth element that makes up the main pillars of this new industrial era. The authors emphasize that CPS are based on manufacturing systems virtualization, aiming to adopt advanced techniques and tools for decision support.

Considering the virtualization of manufacturing processes, Tao and Zhang [4] reveal that the adoption of digital twins (DTs) is becoming increasingly frequent and represents a valuable tool for decision-making support. In this case, we have computational models capable of connecting with physical equipment and systems, mirroring their behavior over time, allowing complex analysis in real/near real-time, and aligned with the current state of production systems [4, 5]. According to Santos et al. [6], DTs of manufacturing processes have become popular in recent years and, in addition to commercial software packages, which usually accompany equipment, DTs based on simulation models stand out, with emphasis on discrete event simulation (DES) models.

The adoption of DTs based on simulation models has been highlighted as a highly flexible and low-cost approach compared to commercial packages, allowing the implementation of DTs in different processes, including those with a low degree of automation [7]. However, although it is a promising approach, there are some challenges regarding the adoption of this approach, with an emphasis on the difficulty of guaranty the model's reliability over time [8]. This fact is critical and can impact

✉ Carlos Henrique dos Santos
chenrique.santoss@gmail.com

[1] Production Engineering and Management Institute, Federal University of Itajubá, Ave. BPS, 1303, Itajubá, Minas Gerais, Brazil

the use of DTs for decision-making since their use is usually associated with high-impact decisions on production processes [9]. Thus, Santos et al. [6] point out that, once the model is connected and synchronized with the physical systems through its data, there may be problems that compromise model results, such as unscheduled stops, machine breakdowns, and communication failures.

On the one hand, there is a gap in the literature regarding works focused on the reliability of DTs, especially when dealing with simulation-based DTs, as highlighted by Zhuang and Liu and Xiong Wright and Davidson [8, 9]. In this case, different from traditional simulation models, which are verified and validated during their building phase, the DT simulation models must also to be periodically evaluated in order to assess their capacity to support decisions. According to Sargent [10], an accreditation phase intends to guarantee that the models are reliable and ready to support decisions and, in this case, we highlight the need for techniques focused on DT simulation models accreditation. On the other hand, evaluating models over time is a difficult task given the complexity of the models, which may involve several parameters to be evaluated, a high level of detail, and real-time/near real-time updates [6, 11]. In this case, we highlight process monitoring as a valuable technique to evaluate processes over time, being a widely disseminated area in the literature in the last decades. Through the so-called control charts, it is possible to monitor complex and stochastic processes in a practical and efficient way, as highlighted by Costa et al. [12].

Thus, we highlight the opportunity to adopt process monitoring principles and techniques in assessing simulation-based DTs to ensure their accreditation in decision support. In this case, the present work proposes a mechanism for monitoring DT models over time, making them more robust and reliable in supporting decisions in production processes. To this end, the models are initially measured as to their correspondence with the physical systems, and, in this case, we used a machine learning technique called "K-Nearest Neighbors" or K-NN [13]. Then, a key indicator resulting from K-NN, which represents the similarity between the physical and virtual environments, is plotted on an attribute control chart [14], the p chart, allowing continuous monitoring of the DT throughout its operation. In this way, we highlight the novelty of this work, which fills a technical and theoretical gap through the combined use of techniques widely adopted by researchers and practitioners.

The rest of the paper is as follows: "Sect. 2" presents the theoretical background, and the proposed approach is described in "Sect. 3, Materials and Methods". "Sect. 4" is dedicated to the application of the proposed approach in two real case studies. Finally, "Sect. 5" concerns the conclusions of the work.

2 Theoretical background

The adoption of simulation models for decision support in production processes is not a recent approach and has been standing out in recent decades as a valuable tool for more efficient decisions [15]. In this case, the simulation can help in decisions involving resource allocation, layout improvement, production planning, staff scheduling, analysis of waste, improvements in the shop floor, and analysis of interactions with other support areas, among others [15]. However, with the growing need for faster decisions and considering the evolution of production processes towards Industry 4.0, the use of simulation has evolved over the last few years towards the DTs era [16].

The Industry 4.0, an allusion to what would be the fourth industrial revolution, is related to more efficient production processes through the adoption of emerging technologies and solutions [1]. In this case, Xu et al. [2], Yin et al. [17], and Moeuf et al. [18] state that solutions such as IoT, Big Data, Cloud technology, CPS, digital twins, artificial intelligence, collaborative robots, and virtual reality have been increasingly frequent in decision support, helping managers in complex decisions. Rüttimann and Stöckli [19] state that Industry 4.0 is an inevitable trend and organizations must prepare for changes in their production structure. In this context, the adoption of simulation models as DTs represents a competitive advantage for decision-makers, since it is a more flexible and cheaper alternative compared to commercial packages and software [20].

The concept of DTs is relatively recent and was initially proposed by the North American Space Agency, NASA, in 2010. Initially planned as virtual copies of physical equipment, the main objective of DTs was to help decision-makers through tools and analysis techniques capable of optimizing decisions while mirroring the physical equipment [21]. Since their creation, DTs have been widely used in the most diverse areas, such as logistics, service, healthcare, and manufacturing [6, 9]. In this case, Tao and Zhang [4] report that a DT has four main elements: (i) physical systems, composed of equipment, machines, processes, among others; (ii) virtual model, composed of a digital copy similar to the physical systems; (iii) systems data, comprising data and information about both physical and virtual environments; and (iv) integration between physical and virtual environments through their data, allowing synchronism between them.

Using simulation as a DT implies several characteristics that differentiate it from traditional approaches. In this case, Davidson [9] reveals that the main difference between a traditional simulation model and a DT model is its ability to update over time according to changes in

physical environments, acting as a virtual mirror of the processes. In addition, the connection and synchronization between the model and the production processes are allowed by collecting and sharing data between the physical and virtual environments through sensors, smart devices, management systems, databases, among others [6]. In this way, when used as a DT, the simulation model becomes a daily tool for decision-making. Santos et al. [20] emphasize that there are approaches from autonomous models that operate in real time to models that update in discrete periods (near real time) and that only suggest decisions, that is, a non-autonomous approach. Figure 1 illustrates the general architecture of the simulation as DT.

Although the adoption of DT simulation models is an increasingly frequent practice for decision support in production processes, there are still challenges to be overcome by modelers, with great emphasis on the need for methods and techniques focused on ensuring the accreditation of models during their operation [6]. Unlike traditional simulation approaches, where the model is evaluated only during its building phase and, once validated, the model is considered valid to support specific decisions and with limited scope [22], the DT model is updated numerous times to mirror the physical systems and, in this case, just traditional validation is not enough. Meng et al. [11] emphasize that ensuring the accreditation of dynamic models that adapt over time is more complex than traditional models.

Santos et al. [6] identified in a systematic literature review that the vast majority of studies do not consider the accreditation of the simulation models during their operation as DTs. In this context, the only approach observed by the authors is the adaptation of traditional validation methods, as proposed by Cho et al. [23], where hypothesis tests are carried out periodically to compare data from the physical and virtual environments. Although statistical techniques such as hypothesis testing are widely used in the validation of traditional simulation models, as highlighted by Sargent and Sargent [22, 24], their use in DT models may not be an efficient approach due to a few reasons: (i) difficulty in obtaining a frequency adequate testing; (ii) the need for different tests simultaneously, considering different types of variables analyzed; (iii) difficulty in analyzing the DT past behavior.

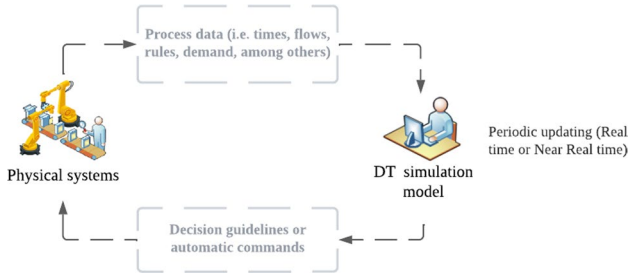

Fig. 1 Architecture of simulation models as DTs

For Tao and Zhang [4], to ensure the DT model accreditation, it is necessary to compare its results with the physical environment periodically. In this case, we highlight the opportunity to evaluate the models through a technique widely used for monitoring processes subject to variability over time, the control charts [25]. According to the authors, the control charts are one of the main statistical process control techniques, helpful in monitoring output variables in systems under undesirable sources of variability. Initially proposed by Walter Shewhart in 1924, the control charts are used to indicate sources of variability and, consequently, the need for corrective actions [26, 27]. Abbas et al. [28] highlight that a control chart is considered efficient if it can quickly notify unusual variations in the process, indicating that the process is not operating as expected.

The study involving the adoption of control charts can be divided into two main stages [27, 29]. Phase I focuses on chart building, choosing the most appropriate chart, and defining its main parameters (upper and lower control limits and center line). On the other hand, phase II focuses on periodic monitoring of the process by measuring key variables we want to monitor, and then plotting this measurement on the chart. To correctly choose the chart to be used, the nature of the monitored variable must be considered. In this case, we highlight the "variable charts" and "attribute charts," which focus on quantitative and qualitative variables, respectively [25].

Considering the use of control charts for the periodic monitoring of DT simulation models, we must create a mechanism to measure how reliable the model is for decision support, that is, to measure the similarity between the model and the physical system behavior. In this case, we highlight the opportunity to use machine learning (ML) techniques to compare physical and virtual environments, emphasizing the K-nearest neighbors (K-NN) technique. According to Lee et al. and Wan et al. [30, 31], the K-NN has stood out in recent years among several ML techniques due to its simplicity and efficiency in solving classification, regression, and clustering problems. Furthermore, considering classification problems, K-NN can classify a dataset (test data) based on other "K" neighboring data already known (training data) and, in this case, the algorithm uses the Euclidean distance to define the data considered the nearest neighbors [13].

Despite its broad applicability, one of the main challenges related to using K-NN is the choice of the "K" parameter, which will vary according to the characteristics of the application. Furthermore, as K-NN is a supervised ML method, the algorithm training step is directly linked to the quality of its results [32, 33]. Finally, according to Kumar et al. [13], the accuracy of K-NN is a metric that indicates how well the data can be classified, that is, considering two sets of data,

an accuracy of 100% indicates that the classifier can identify the difference in the data and classify them correctly. On the other hand, if the data are perfectly similar, the classifier will not be able to identify the differences and assign a random classification, that is, an accuracy of 50%. Thus, it is possible to measure the similarity between the DT model and the physical systems through the accuracy of the K-NN. In other words, data from the physical and virtual systems will be compared and classified through the K-NN, whose accuracy could be plotted on a control chart.

3 Materials and methods

Considering the objective of creating a monitoring tool capable of evaluating DT simulation models over time and ensuring their accreditation to support decisions, the present work is based on two main steps:

(i) Tool-building phase: in which all the necessary structure for the operation of the monitoring tool is planned and developed
(ii) Tool operation phase: in which there is the periodic collection of data, the comparison between the physical and virtual environments through K-NN, and the control chart updating

3.1 Tool-building phase

The tool building was planned to allow the decision-maker to have a friendly and easy-to-use interface. In this case, the algorithm was developed in Python, using the Scikit-learn library [34]. The tool must be able to carry out several activities, as described below.

First, we highlight that the tool must be able to read data from both physical and virtual environments. Analogously to traditional validation methods, the data to be compared must be the critical information to the model proposal, which we will call "evaluation variables" [22]. The choice of evaluation variables will depend on the characteristics of each application, and the tool must be able to collect such variables in the same proportion of the physical and virtual environment. Some typical evaluation variables in manufacturing systems are process times, waiting times for activities, process and equipment parameters, and resource utilization rates. Furthermore, since the evaluation variables can be of different types (e.g., numeric and categorical, integer and real), the algorithm must also scale the read data and, in this case, we adopt the Robust Scaler method, which scales the data according to the interquartile range [35].

Once collected and scaled, the evaluation variables are compared through the K-NN classifier. In this case, the K-NN will try to differentiate the data from the physical and virtual environments. Suppose the DT simulation model is operating as planned and its results are similar to the behavior of physical systems. In this case, the K-NN will not be able to classify them and will present an accuracy of around 50%. Thus, an accuracy of 50% of the K-NN would indicate a valid and reliable DT simulation model. In comparison, an accuracy far from 50% indicates an invalid model with low reliability results, in other words, one or more evaluation variables from the digital model differ significantly from those of physical environment. For the use of K-NN, the correct choice of the parameter "K" and the training of the algorithm are fundamental steps, and we adopted the "cross-validation" for these purposes. In this case, the training data is divided into "Z" folds which will be trained and tested in "Z" times [36, 37]. As suggested by Li et al. [38], we used 5-fold cross-validation. Furthermore, several "K" values are tested during the training phase and the algorithm will use the one that results in the best classifier accuracy.

Once the accuracy of the K-NN is obtained, we can consider that the similarity between the model and the physical systems has been measured. Therefore, the next step is to plot this measurement on the control chart aiming at its monitoring over time. In this case, as the accuracy of the K-NN is obtained from the proportion of successes and errors of the classifier, we adopted an attribute control chart, the *p* chart. Attribute charts are widely used by researchers and decision-makers due to their simplicity and practicality. In this case, we highlight the *p* chart, which allows monitoring the proportion of observations with a particular characteristic of interest [14, 29]. Therefore, the monitoring tool was configured to execute the two phases of the control chart. First, the parameters of the control chart are defined and then the graph is used periodically to monitor the DT simulation model.

Considering the control chart *p* building, Abbas et al. [28] emphasize that the central line (CL) refers to the proportion "*p*" of an evaluated characteristic and, in the case of this application, the value of "*p*" was adopted as 0.5, since it refers to the K-NN accuracy target. In addition, the lower (LCL) and upper (UCL) control limits are obtained through eqs. (1) and (2).

$$\mathrm{UCL} = p + 3\sqrt{\frac{p(1-p)}{n}} \tag{1}$$

$$\mathrm{LCL} = p - 3\sqrt{\frac{p(1-p)}{n}} \tag{2}$$

In which *n* represents the sample size adopted. In this case, the algorithm will accumulate the first 25 observations (sample size of "*n*") to estimate the UCL and LCL values, as suggested by Montgomery [25]. It is essential to highlight that, during this phase I of the control chart study, we must

ensure that the model and the physical systems are validated and operating as planned, without special causes that may compromise the results. The sample size "n" can be obtained through Eq. (3).

$$n = \left(\frac{L}{\delta}\right)^2 p(1-p) \tag{3}$$

L is the distance of the control limits from the CL and δ is the magnitude of the process shift. If we consider that $\widehat{p}_0$ is the estimated in-control value of p and $\widehat{p}_1$ is the out-of-control value, then $\widehat{p}_1 = \widehat{p}_0 + \delta$.

In this case, we considered a fraction "p" of 0.5, the parameter L of 3 standard deviations, and the magnitude of the process shift "δ" of 0.15 (to guarantee that the accuracy of the K-NN will vary around 0.5 ± 0.15, which according to our proposal will guarantee satisfactory accreditation of the DT simulation model). Therefore, we obtained "n" = 100. In other words, the monitoring tool was configured to read samples of 50 from both environments (physical and virtual), totaling a final sample of 100. Furthermore, once we defined the control chart parameters (phase I), we started phase II of the control chart, the operation phase.

3.2 Tool operation phase

The operation phase of the monitoring tool is associated with its use by decision-makers to obtain more reliable and efficient DT simulation models. Thus, this work aims to provide a tool for periodic use for better decisions, and, in this case, the monitoring frequency is an important characteristic. Considering that current production systems have an increasingly efficient and practical data collection structure, we note that data from physical and virtual systems (evaluation variables) are continuously collected overtime. Furthermore, once the sample size is reached, the tool will be updated, and, consequently, the DT simulation model will be monitored. Finally, suppose the control chart indicates an out-of-control point. In that case, we suggest that the modelers and decision-makers reevaluate the DT structure (including the model and the physical systems) in order to correct possible problems such as communication failures between the model and the physical equipment, model logic problems, hardware and/or software failures, unscheduled process stops, among others. Figure 2 illustrates the monitoring tool structure as presented before and Fig. 3 summarizes the activities carried out in the building and operation phases of the tool.

4 Results and discussions

In order to evaluate the applicability of the proposed approach, we implemented it in real study objects. In this case, we chose two simulation-based DTs that are already in the operational phase, which were built and ready to support decision-making. It is worth mentioning that this work does not address issues related to the DTs building, as proposed by [39], and, therefore, we assume that the DTs are operational and we can implement the monitoring tool as a supplement to the DTs, aiming at their accreditation in order to guarantee greater robustness and reliability of their results.

4.1 Study objects

The case study I refers to a medium-sized production line that produces clothing items. It is a production process inserted into the fast-fashion segment, resulting in demands with high variety and volume and which lacks faster and more accurate decisions. The line has eight workstations (processes B to I), in addition to reception (process A) and shipping (process J), which produce three products with

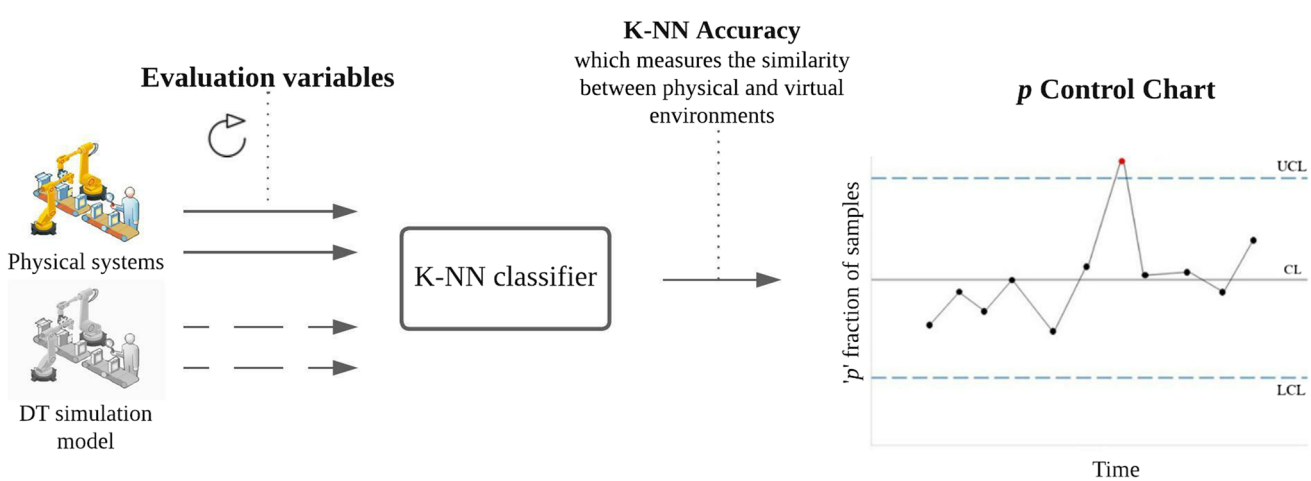

Fig. 2 Monitoring tool architecture

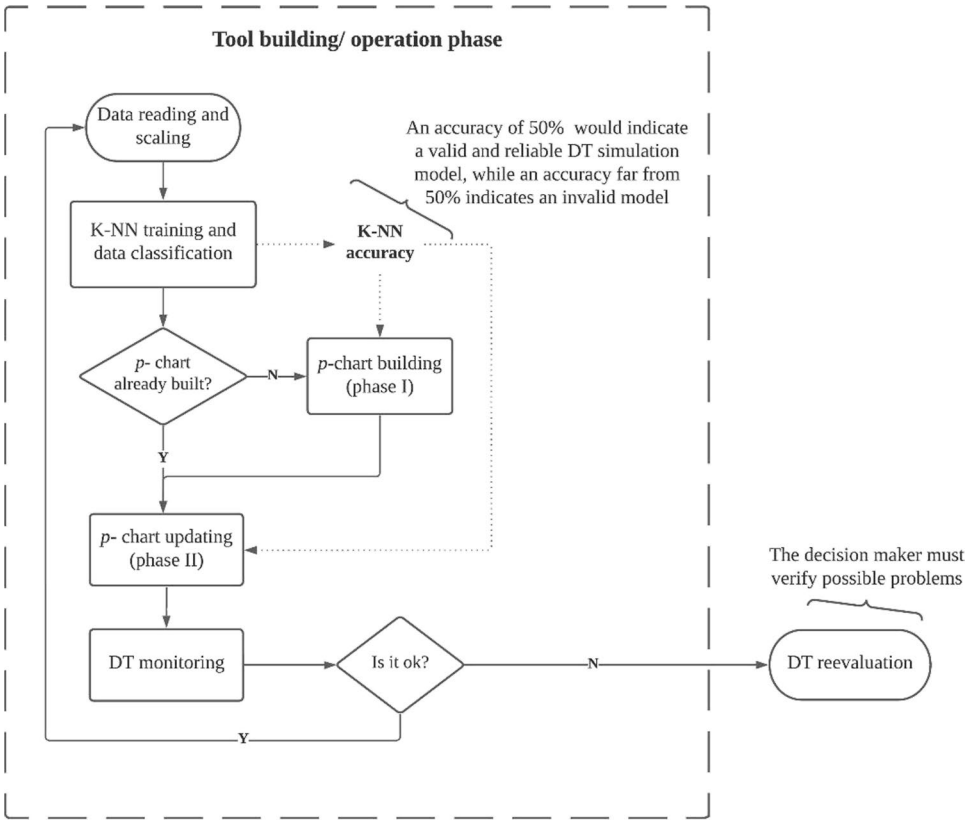

Fig. 3 Activities carried out in the building and operation phases of the tool

different behavior of demands. All workstations are manual, and only the transport of products and materials between them is carried out by an automated guided vehicle (AGV). In this case, the DT was designed to support operational decisions and was based on a DES model built using FlexSim®. It is important to highlight that the DT was previously built, verified, and validated by the team responsible for the project and is already operational. In other words, the decision-makers are already using DT to support decisions.

The DT operates in near real time, and it is a non-autonomous approach. In this case, the model is updated weekly and provides guidelines for decision-making. In addition to the DES model, the DT is based on an artificial intelligence (AI) tool and a decision dashboard. Basically, data related to the historical product demand is collected and inputs the AI algorithm, based on artificial neural networks (ANN), which predicts the weekly demand behavior. Furthermore, the DES model reads the AI output and tests different resource planning strategies, resulting in the best decision regarding resource sizing (physical and human). The decision dashboard has an essential role in the system since it integrates the physical and virtual environments and provides a user-friendly interface for the decision-maker. Through automated buttons, the user can run the AI and DES models and view the decision-making guidelines. More details about the case study I can be found in Santos et al. [40], and Fig. 4 illustrates the production flow, the DES model, and the DT architecture described above.

The DES model was built in FlexSim® due to some important features that are essential when we consider simulation-based DTs, such as the visual representation, the integration of the model with optimization techniques, the possibility to use virtual reality, and the ease of connection with the local databases and management systems. Moreover, since the DT is used to support operational decisions, the model updating depends on the processes and products status and states. Therefore, we defined each product's processes times and waiting times as critical parameters that must be considered when selecting the evaluation variables. Finally, it is important to highlight that other process parameters were not considered critical for the DT operation since they do not impact the model updating or can be obtained through the information already collected. Thus, the objective of the monitoring tool is to compare the processes times and waiting times from both virtual and physical environments to allow the DT accreditation assessment.

On the other hand, the case study II is an automated production cell. When considering an automated production line, it is important to guarantee the correct functioning of all the equipment and systems. In this context, the selected DT is also based on a DES model and focuses on monitoring a production cell. This cell has a simple structure, with

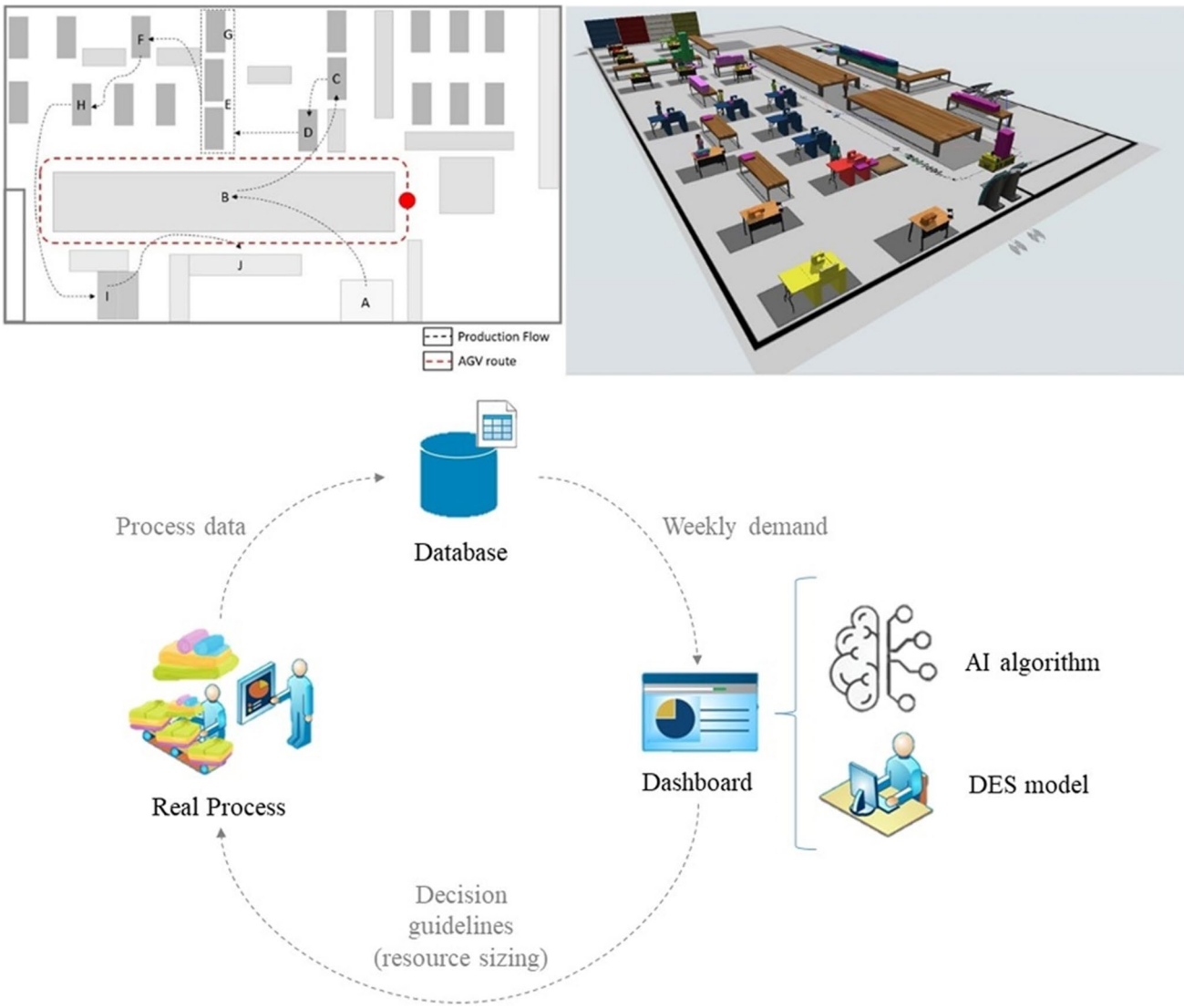

Fig. 4 Production flow, DES model, and DT architecture (case study I)

two workstations that manufacture two types of products, considering three working shifts. First, the raw materials arrive in the arrival area (A) and are sent to the supply areas (B and D), which supply the raw material to the workstations (C and E), respectively. An automatic conveyor carries out the transport between the areas, and the movement between these areas and the workstations is carried out by robotic arms (R1 and R2). The workstations C and E are lathe and milling CNC machines, respectively.

According to Santo et al. [6], the use of virtual models to monitor production systems and evaluate metrics during their operation is one of the main applications of simulation-based DTs. In this case, the DT was planned to support the productive cell through key performance indicators (KPIs), such as total lead time, stop times, production rate, and overall equipment effectiveness. The process data are collected in real time through sensors and smart systems and stored in the local database. The current work in progress (WIP) is used to update the DES model in near real time with a few minutes delay, mirroring the cell in a virtual environment and allowing the comparison between the expected and the real behavior. In this case, we have a non-autonomous approach since the DT is used to evaluate the process but not to infer about it. A decision dashboard is also adopted to allow the integration between the DES model and the physical systems, and the DT was previously built, verified, and validated. Figure 5 illustrates the production flow, the DES model, and the DT architecture.

In the same way as case study I, the DES model was built in FlexSim® due to the software features that are in line with DT requirements. Furthermore, we reinforce that the DT was planned to support the process assessment

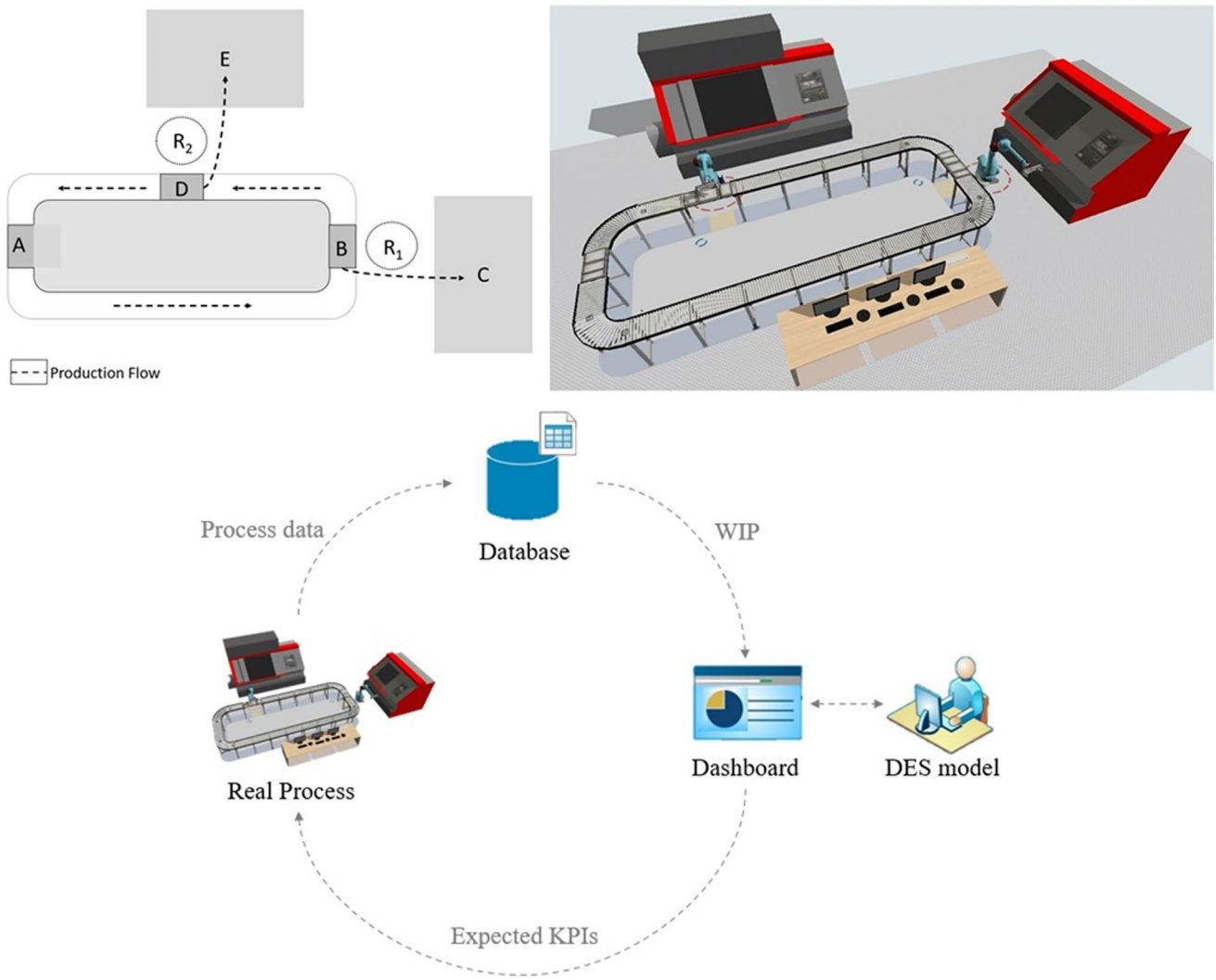

Fig. 5 Production flow, DES model and DT architecture (case study II)

instead machines assessment. Therefore, the processes and products states and status are critical parameters that impact the DT operation. In this case, the evaluation variables are also based on each product's processes times and waiting times. Thus, the monitoring tool should compare this information from virtual and physical environments to assess the DT accreditation. It is important to highlight that the machines' parameters were not considered for the DT monitoring since they do not impact model updating. In this sense, each machine was modeled as a process with input and output rates and that processes products according to a stochastic process time. This information was based on the optimal parameters set in the physical systems and which do not change often. In this case, if these machines' parameters change, process times and rates can be updated manually in the model. Finally, analyses focused on the machines' assessment can be considered in future studies.

4.2 Experimental results

4.2.1 Case study I

For case study I, the chosen evaluation variables were each product's waiting times and process times, considering the main workstations (processes C to I) and the travel time by AGV (totaling 14 variables). In this case, this product information is periodically collected through radio frequency tags (RFID) and stored in the local database. Moreover, the DES model also periodically collects the same information from the virtual environment. Therefore, virtual and physical data are stored in the same proportions in different datasets, ready to input the monitoring tool.

It is important to highlight that the average weekly demand is around 750 clothing items (including all product types). Therefore, the monitoring tool was configured to collect the physical and virtual data in samples of 50

during the week, totaling datasets of 100 (sample size defined in "Sect. 3"). Thus, if the line produces 750 items weekly, about 15 samples will be collected. We considered one month of DT operation for this work to evaluate the monitoring tool. Then, all monitoring tool activities were carried out at each sampling, including the data collection and scaling, the K-NN training and classification, and the control chart building and updating. As defined in "Sect. 3," the first 25 observations were used to define the control chart parameters, while the other observations were used to update the control chart. Figure 6 presents the control chart for case study I.

We must highlight that we did not induce any special causes in the DT during the experiments. Therefore, we note that the monitoring tool works as expected, considering its use in a real case study. The K-NN average accuracy represented a reliable and validated DT during the assessed period, and, despite the complexity of the system (14 evaluation variables), the monitoring tool was able to compare the physical and real environments and identify their similarities. Furthermore, it is also possible to evaluate the control chart performance and, in this case, Leoni and Costa [14] and Aebtarm and Bouguila [41] suggest some assessment indicators: (i) plotting illustration and (ii) average run length (ARL). These indicators refer to the visual quality of the chart and the rate of false alarms, respectively. The *p* chart ARL can be obtained by Eq. 4.

$$\mathrm{ARL} = \frac{1}{pr} \tag{4}$$

In which pr represents the probability that a point is out-of-control (considering the Binomial distribution).

Considering the plotting illustration, Aebtarm and Bouguila [41] highlight that the control chart should clearly present the evaluated system's behavior, allowing to identify whether something is right or wrong. As shown in Fig. 6, the monitoring tool fulfills this role. On the other hand, considering the ARL, Leoni and Costa [14] state that a larger ARL means a lower false alarm rate and, for case study I, we obtained an ARL of 714, considering a pr of 0.0014 (obtained from binomial distribution and considering CL = 0.61, UCL = 0.76, and LCL = 0.46). This result means that an out-of-control point will appear at every 714 observations if the process is in control. Aebtarm and Bouguila [41] reveal that an ARL of 370 is commonly used as a reference value considering three standard deviation levels for the chart limits and, therefore, the application of the proposed tool in the first case study was successful.

4.2.2 Case study II

Considering the selection of evaluation variables of case study II, we also opted for each product's waiting times and process times, considering both workstations C and E (totaling four variables). Sensors continuously collect these variables and smart systems (physical data), and the DES model periodically offers the same information from the DT (virtual data). Physical and virtual data are also periodically stored in different datasets, ready for monitoring. The average shift demand is around 110 items (including the two product types), and the monitoring tool was planned to collect the physical and virtual data in samples of 50 during the working day, totaling datasets of 100. Moreover, we recorded observations from about 50 working shifts, and, as previously defined, the first 25 observations were used to define the limits of the control chart, while the others were used in the monitoring process. Figure 7 shows the control chart considering the described period.

In the same way as in case I, the monitoring tool works as expected. We note that the tool was able to compare the physical and real environments and identify their similarity, as indicated by the K-NN average accuracy. Since we did not induce any special cause in the process, we observed a reliable and validated DT in the assessed period, as expected. First, we highlight that the control chart presents the behavior of the evaluated system. Furthermore, regarding the false alarm rate, we obtained an ARL of about 625, considering a pr of 0.0016 (obtained from binomial distribution and

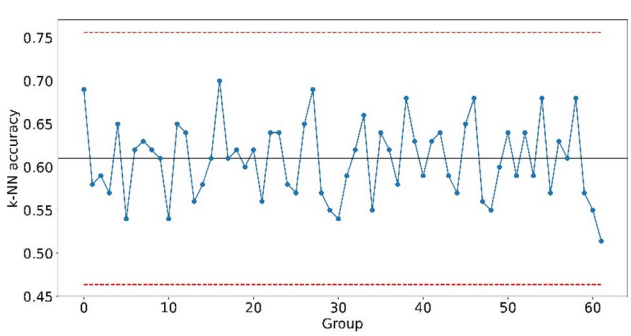

Fig. 6 Case study I control chart

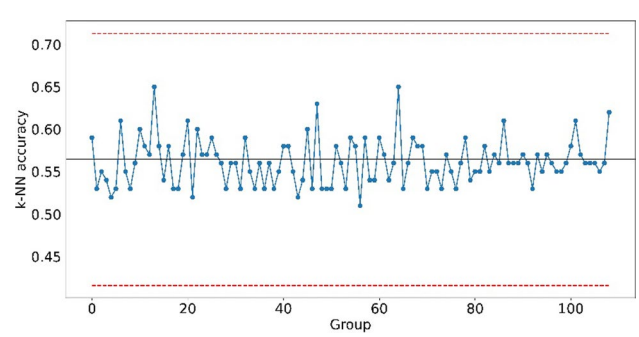

Fig. 7 Case study II control chart

considering CL = 0.57, UCL = 0.72, and LCL = 0.42). In other words, an out-of-control point will appear at every 625 observations if the process is in control. Therefore, we can conclude that applying the proposed tool in the second case study was also successful.

4.2.3 Final discussions

Considering the experimental results, we note that the proposed approach simplifies the DT assessment. It is possible to analyze several variables with only one parameter (K-NN accuracy), plotted in one control chart. This feature is especially valid when considering complex DTs with several evaluation variables. Compared to traditional approaches based on hypothesis tests, each evaluation variable should be compared in pairs, resulting in a longer monitoring and limiting if the DT model is updated in short time intervals. For example, case studies I and II would require about 14 and 4 simultaneous and periodic hypothesis tests, respectively. Moreover, there may be a need to adopt different tests depending on the nature of the evaluation variables (i.e., parametric and non-parametric tests). Finally, considering multiple hypothesis tests, the confidence level might be harmed since there is a combined effect of the error associated with each test. In contrast, the proposed monitoring approach integrates reliability and agility in monitoring simulation-based DT models.

Furthermore, we carried out a comparative analysis to demonstrate the advantages of the proposed approach compared to current practices based on periodic hypothesis tests. In this case, we considered the same monitoring period described before: 1 month of operation for case study I and 50 working shifts for case study II. First, we verified the normality of the data to choose the proper hypothesis test and they did not fit with the normal distribution (*P*-value of the Anderson–Darling test < 0.05). Then, at each collected sample (composed of data from physical and virtual environments), we performed the Mann–Whitney test for each evaluation variable to compare the physical and virtual data. In this case, if the DT is working as expected, the test should present a *P*-value ≥ 0.05. Figure 8 illustrates the assessment of case studies I and II through this approach.

We note that, although it was possible to evaluate the DTs over time through hypothesis tests, this approach may be impractical or unfeasible. For case study I, it was necessary about 840 hypothesis tests to evaluate the DT during 1 month of operation and considering all 14 evaluation variables. In the same way, for case study II, about 400 hypothesis tests were necessary to allow the assessment of the DT during 50 working shifts and considering all 4 evaluation variables. As highlighted before, this approach can become even more complicated if there is a need for different hypothesis tests or the inclusion of more evaluation variables. On the other hand, the proposed approach allows the evaluation of the DTs in a simpler and faster way, as demonstrated in this work.

In addition, considering the correlation between the evaluation variables, the monitoring would require

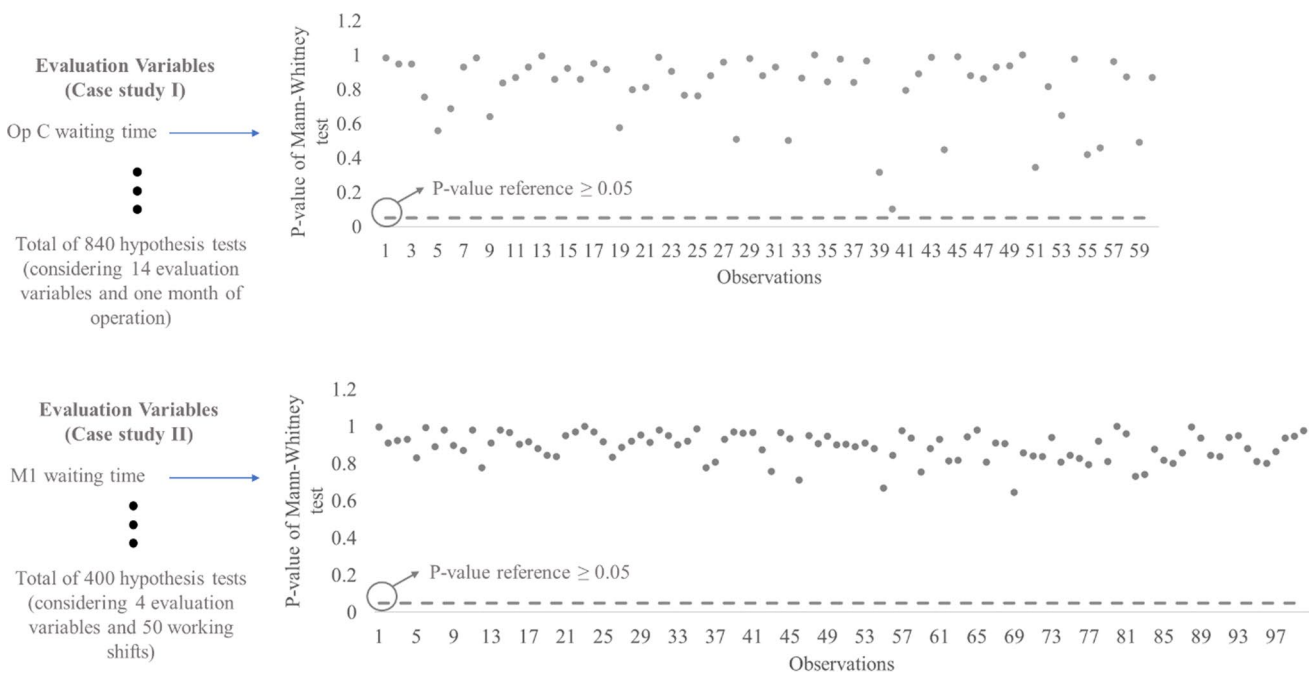

Fig. 8 Case studies I and II assessment through periodical hypothesis tests

additional analysis. Thus, if a particular variable presents an unexpected behavior (a result of a special cause), the other correlated variables will be impacted, resulting in wrong conclusions about the DT accreditation. In contrast, the proposed approach is not based on individual analyses but on monitoring all variables in an integrated manner, making monitoring more agile. Furthermore, the monitoring tool may present the correlation matrix of the evaluation variables, indicating to the decision-maker the levels of correlation between the variables and facilitating the investigation of possible instances of special causes if the process is out-of-control. It is verified in case study I DT, which has correlated evaluation variables, as shown in Fig. 9. In this case, we note that the "C process time" and "D waiting time" variables have a strong correlation.

5 Conclusion

The simulation-based digital twins have become widely adopted by researchers and professionals aiming at increasingly efficient decision systems aligned with the behavior of physical systems. Furthermore, it is an approach aligned with the principles and pillars of the Industry 4.0, representing an inevitable trend for production systems. However, as with all new developments and solutions, there are challenges associated with adopting DTs in decision support, with emphasis on the accreditation of the models over time. In this case, since the use of DTs is usually associated with high-impact decisions on production systems, ensuring their accreditation during operation is an important requirement. There is a gap in the literature regarding works focused on methods and techniques for this purpose.

Therefore, this work proposed a tool focused on monitoring simulation-based DTs to assess their accreditation during the operational phase, that is, during their use in decision support. Different from the validation steps during the traditional simulation model building, the proposed approach focuses on a periodical model assessment in order to guarantee the reliability of their results overtime. First, to measure the correspondence between the physical and virtual environments, we adopted the K-NN classifier, which compares data from both environments and provides a metric that represents how well the model represents the physical systems. In addition, this metric is monitored over time through the *p* control chart, a technique widely used in monitoring processes subject to variability.

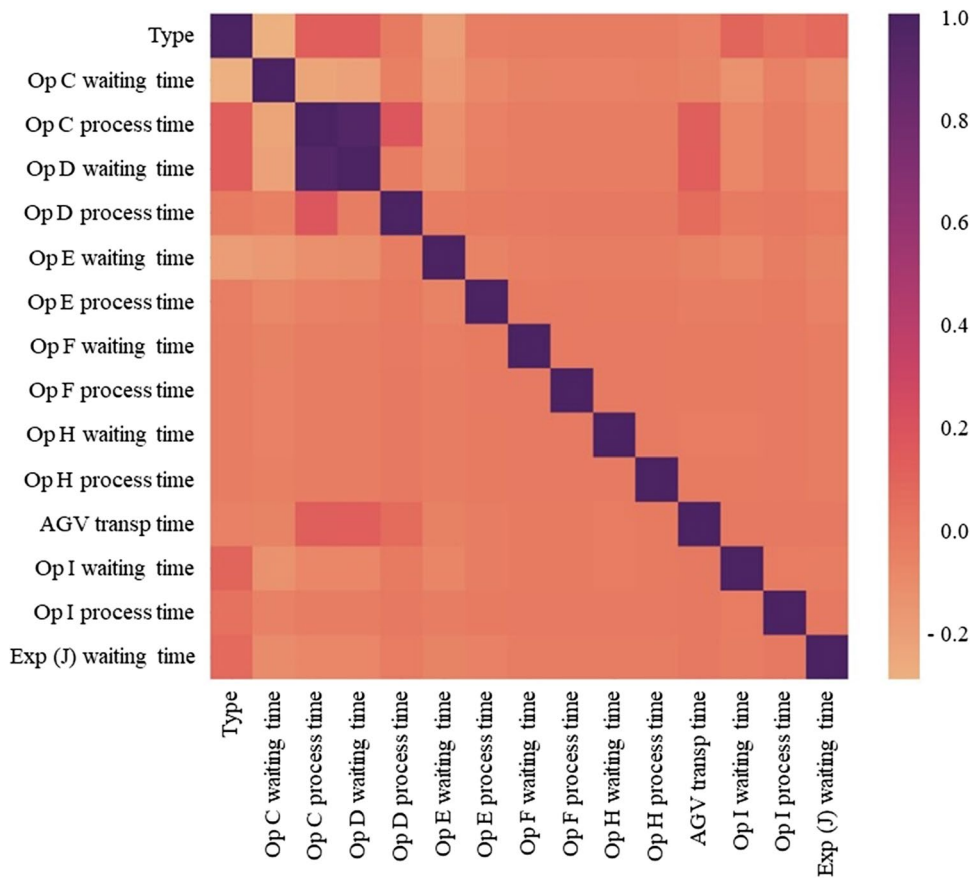

Fig. 9 Correlation matrix between evaluation variables (case study I)

The proposed tool was coded in Python and provides an end-user-friendly interface. In this case, the tool periodically collects data from both physical and virtual environments, scales them, compares them using the K-NN, and, finally, updates the control chart from the K-NN results. In this way, the user can assess the accreditation of the DTs simulation models. The proposed approach was applied to two real objects of study, which refer to DTs already implemented and in the operational phase. The first object, case study I, refers to a DT focused on supporting the resource planning of a manual production line that belongs to the fast-fashion segment. On the other hand, the second object, case study II, refers to a DT focused on mirroring an automated production cell. In this case, the model was planned to evaluate the main process metrics. As a result, we observed that the tool proved to be efficient in monitoring both systems, making the DTs more reliable. It was possible to monitor several evaluation variables simultaneously and in a simpler and faster way if we compare the proposed tool with traditional approaches of validation, such as hypothesis tests.

The tool proved to be agile in monitoring DTs and independent of the system characteristics, that is, it can be adopted in autonomous models that operate in real-time and near real-time models that only suggest decisions and do not control the physical equipment. Finally, we noted that the tool could monitor complex systems with DTs that involve several evaluation variables of different types (e.g., numerical and categorical, integer and real, among other combinations). Finally, for future work, we suggest its application in other objects of study, including DT models that are not based on simulation. In addition, it is also suggested the evaluation of other ML techniques and other control charts in order to improve the tool and provide comparative analysis. Finally, we suggest improvements in the tool to identify and indicate the impact of each evaluation variable on DT accreditation.

Acknowledgements The authors thank CAPES, CNPq, FAPEMIG, and UNIFEI Pos-graduate Program.

Funding This work was funded by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Uriarte AG, Ng AHC, Moris MU (2018) Supporting the lean journey with simulation and optimization in the context of Industry 4.0. Procedia Manuf 25:586–593. https://doi.org/10.1016/j.promfg.2018.06.097
2. Da XuL, Xu EL, Li L (2018) Industry 4.0: state of the art and future trends. Int J Prod Res 56:2941–2962. https://doi.org/10.1080/00207543.2018.1444806
3. Zhong RY, Xu X, Klotz E, Newman ST (2017) Intelligent manufacturing in the context of industry 4.0: a review. Engineering 3:616–630. https://doi.org/10.1016/J.ENG.2017.05.015
4. Tao F, Zhang M (2017) Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing. IEEE Access 5:20418–20427. https://doi.org/10.1109/ACCESS.2017.2756069
5. Alam KM, EL Saddik A (2017) C2PS : a digital twin architecture reference model for the cloud-based cyber-physical systems. IEEE Access 5:2050–2062. https://doi.org/10.1109/ACCESS.2017.2657006
6. dos Santos CH, Montevechi JAB, de Queiroz JA et al (2022) Decision support in productive processes through DES and ABS in the digital twin era : a systematic literature review. Int J Prod Res 60:2662–2681. https://doi.org/10.1080/00207543.2021.1898691
7. Santos CH, Queiroz JA, Leal F, Montevechi JAB (2022) Use of simulation in the industry 4.0 context: creation of a digital twin to optimize decision making on non-automated process. J Simul 16:284–297. https://doi.org/10.1080/17477778.2020.1811172
8. Zhuang C, Liu J, Xiong H (2018) Digital twin-based smart production management and control framework for the complex product assembly shop-floor. Int J Adv Manuf Technol 96:1149–1163. https://doi.org/10.1007/s00170-018-1617-6
9. Wright L, Davidson S (2020) How to tell the difference between a model and a digital twin. Adv Model Simul Eng Sci 7:1–13. https://doi.org/10.1186/s40323-020-00147-4
10. Sargent RG (2011) Verification and validation of simulation models: an advanced tutorial. In: Proceedings of the 2011 Winter Simulation Conference. pp 183–198
11. Meng C, Nageshwaraniyer SS, Maghsoudi A et al (2013) Data-driven modeling and simulation framework for material handling systems in coal mines. Comput Ind Eng 64:766–779. https://doi.org/10.1016/j.cie.2012.12.017
12. Costa AFB, De MMS, Epprecht EK (2009) Monitoring the process mean and variance using a synthetic control chart with two-stage testing. Int J Prod Res 47:5067–5086. https://doi.org/10.1080/00207540802047098
13. Kumar RP, Melcher D, Buttolo P, Jia Y (2020) Monitoring of occupant states in autonomous vehicles using capacitance-sensing imaging. IEEE Sens J 20:12914–12927. https://doi.org/10.1109/JSEN.2020.3003312
14. Leoni RC, Costa AFB (2018) The Shewhart attribute chart with alternated charting statistics to monitor bivariate and trivariate mean vectors. Comput Ind Eng 122:273–282. https://doi.org/10.1016/j.cie.2018.06.003
15. Mourtzis D (2020) Simulation in the design and operation of manufacturing systems: state of the art and new trends. Int J Prod Res 58:1927–1949. https://doi.org/10.1080/00207543.2019.1636321
16. Rodič B (2017) Industry 4.0 and the new simulation modelling paradigm. Organizacija 50:193–207. https://doi.org/10.1515/orga-2017-0017
17. Yin Y, Stecke KE, Li D (2018) The evolution of production systems from Industry 2.0 through Industry 4.0. Int J Prod Res 56:848–861. https://doi.org/10.1080/00207543.2017.1403664
18. Moeuf A, Pellerin R, Lamouri S et al (2018) The industrial management of SMEs in the era of Industry 4.0. Int J Prod Res 56:1118–1136. https://doi.org/10.1080/00207543.2017.1372647
19. Rüttimann BG, Stöckli MT (2016) Lean and Industry 4.0—twins, partners, or contenders? A due clarification regarding the supposed clash of two production systems. J Serv Sci Manag 09:485–500. https://doi.org/10.4236/jssm.2016.96051
20. dos Santos CH, De QJA, Leal F, Montevechi JAB (2022) Use of simulation in the industry 4.0 context: creation of a digital twin

to optimise decision making on non-automated process. J Simul 16:284–297. https://doi.org/10.1080/17477778.2020.1811172
21. Shafto M, Conroy M, Doyle R et al (2010) DRAFT modeling , simulation, information technology & processing roadmap. In: Technology Area 11 - National Aeronautics and Space Administration (NASA). pp 1–27
22. Sargent RG (2013) Verification and validation of simulation models. J Simul 7:12–24. https://doi.org/10.1057/jos.2012.20
23. Cho M, Song M, Yoo S, Reijers HA (2019) An evidence-based decision support framework for clinician medical scheduling. IEEE Access 7:15239–15249. https://doi.org/10.1109/ACCESS.2019.2894116
24. Sargent RG (2014) An interval statistical procedure for use in validation of simulation models. J Simul 9:1–6. https://doi.org/10.1057/jos.2014.30
25. Montgomery DC (2009) Introduction to statistical quality control
26. Apsemidis A, Psarakis S, Moguerza JM (2020) A review of machine learning kernel methods in statistical process monitoring. Comput Ind Eng 142:1–12. https://doi.org/10.1016/j.cie.2020.106376
27. Zwetsloot IM, Woodall WH (2021) A review of some sampling and aggregation strategies for basic statistical process monitoring. J Qual Technol 53:1–16. https://doi.org/10.1080/00224065.2019.1611354
28. Abbas Z, Nazir HZ, Akhtar N et al (2019) On designing an efficient control chart to monitor fraction nonconforming. Qual Reliab Eng 36:1–18. https://doi.org/10.1002/qre.2590
29. Chukhrova N, Johannssen A (2019) Hypergeometric p -chart with dynamic probability control limits for monitoring processes with variable sample and population sizes. Comput Ind Eng 136:681–701. https://doi.org/10.1016/j.cie.2019.07.056
30. Lee C, Huang K, Shen Y, Lee Y (2020) Improved weighted k-nearest neighbor based on PSO for wind power system state recognition. Energies 13:1–16. https://doi.org/10.3390/en13205520
31. Wang MC, Tsai CF, Lin WC (2021) Towards missing electric power data imputation for energy management systems. Expert Syst Appl 174:1–9. https://doi.org/10.1016/j.eswa.2021.114743
32. Hassanat AB, Abbadi MA, Alhasanat AA (2014) Solving the problem of the k parameter in the KNN classifier using an ensemble learning approach. Int J Comput Sci Inf Secur 12:33–39
33. Wang Z, Liao X, Wang C et al (2020) Driver behavior modeling using game engine and real vehicle: a learning-based approach. IEEE Trans Intell Veh 8858:1–12. https://doi.org/10.1109/TIV.2020.2991948
34. Pedregosa F, Varoquaux G, Gramfort A et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830. https://doi.org/10.1289/EHP4713
35. Raju VNG, Lakshmi KP, Jain VM et al (2020) Study the influence of normalization/transformation process on the accuracy of supervised classification. In: Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020. pp 729–735
36. Boontasri K, Temdee P (2020) Determining significant classification factors for senior learning: a case study of thai seniors and social media skill learning. Wirel Pers Commun 115:2951–2970. https://doi.org/10.1007/s11277-020-07216-2
37. Zhang J, Lian J, Yi Z et al (2021) High-accuracy guide star catalogue generation with a machine learning classification algorithm. Sensors 21:1–16. https://doi.org/10.3390/s21082647
38. Li G, Hwai-yong Tan M, Hui Ng S (2019) Metamodel-based optimization of stochastic computer models for engineering design under uncertain objective function. IISE Trans 51:517–530. https://doi.org/10.1080/24725854.2018.1504355
39. Montevechi JAB, Santos CH, Gabriel GT et al (2020) A method proposal for conducting simulation projects in industry 4.0: a cyber-physical system in an aeronautical industry. In: Proceeding of the 2020 Winter Simulation Conference. Orlando, USA, pp 2731–2742
40. dos Santos CH, Gabriel GT, Amaral JVS, do, et al (2021) Decision-making in a fast fashion company in the Industry 4. 0 era: a digital twin proposal to support operational planning. Int J Adv Manuf Technol 116:1653–1666
41. Aebtarm S, Bouguila N (2011) An empirical evaluation of attribute control charts for monitoring defects. Expert Syst Appl 38:7869–7880. https://doi.org/10.1016/j.eswa.2010.12.093