Int j simul model 25 (2026) 2, 236-247
Review scientific paper

CONTINUOUS VALIDATION OF DIGITAL TWINS: CONCEPTUAL AND APPLIED APPROACHES

Lucio, J. S.; Miranda, R. C.#; Montevechi, J. A. B. & Lucio, T. A. V.
Institute of Production Engineering and Management, Universidade Federal de Itajubá (UNIFEI), Itajubá, Brazil
E-Mail: serafim.epr@gmail.com, rafael.miranda@unifei.edu.br, montevechi@unifei.edu.br, thaynaengmec@gmail.com (# Corresponding author)

Abstract
The growing adoption of Digital Twins (DTs) in Industry 4.0 (I4.0) raises the challenge of ensuring model validity throughout the operational lifecycle. Unlike static validation, which is limited to the development phase, continuous validation aims to ensure that the DT remains representative of the physical system during operation. This paper presents a critical review of conceptual and applied approaches to continuous or recurrent DT validation. The studies are classified into continuous monitoring proposals, methods with potential for recurrent adaptation, and conceptually aligned approaches. For each group, methods, results, and limitations are discussed. A comparative analysis reveals relevant gaps, including the lack of formal criteria for selecting critical variables and difficulties in diagnosing the causes of deviations. Finally, emerging trends and future directions are outlined, highlighting the integration of statistical techniques, such as control charts, to enhance the reliability and traceability of DTs in dynamic industrial environments.
(Received in January 2026, accepted in April 2026. This paper was with the authors 1 week for 1 revision.)

Key Words: Digital Twins, Continuous Validation, Industry 4.0, Model Monitoring

1. INTRODUCTION

The digital transformation driven by I4.0 has promoted increasing integration between physical and digital systems, expanding the use of DTs as tools for decision support and process optimization. Technologies such as the Internet of Things (IoT), Big Data, and cloud computing enable the continuous collection and integration of data across devices and models, supporting real-time analysis [1, 2]. In this context, DTs have emerged as virtual replicas of real systems with potential for predictive simulation and real-time monitoring [3].

As DTs assume a central role in industrial operations, making it increasingly important to ensure that these models remain valid over time. Historically, Verification, Validation, and Accreditation (V&V) efforts for simulation models and DTs have focused on the initial development and calibration phases, with limited attention to validity during operation [4, 5]. A review of more than 70 simulation-based DT applications [4] found that only one study included some form of periodic evaluation after implementation, highlighting a significant gap in the literature.

Even among these few examples, most relied on sporadic statistical tests rather than continuous monitoring processes. This gap is critical because models validated only at the outset may become obsolete due to changes in the physical process, sensor failures, or undetected operational drift [6, 7]. Sargent [6] already warned that the performance of simulation models may degrade over time when validation is treated as a one-time activity.

In recent years, several studies have highlighted the need for continuous DT validation. Recent systematic reviews reinforce that most DT research neglects a full lifecycle perspective and focuses almost exclusively on initial model development [8]. Only a small number of studies adopt active revalidation strategies during DT operation. Bitencourt et al. [8] reported that a minimal fraction of publications includes automated continuous validation processes, while most rely on manual inspections or static indicators. Similarly, Biazen et al. [9] pointed out that, even in well-established simulation domains, methods capable of adapting to dynamic systems remain rare, revealing a lack of recurrent or adaptive approaches.

The absence of consolidated lifecycle validation frameworks compromises DT traceability and reliability [10]. In complex industrial settings, where multiple variables are monitored simultaneously, relying solely on initial validation or occasional manual checks is impractical. This leads to the following research question: how can the validity of simulation-based DTs be continuously monitored while ensuring statistically consistent criteria for deviation detection and cause identification?

This paper addresses this question through a systematic literature review focused on continuous DT validation strategies. Without proposing a new method, the study aims to map and critically analyse existing conceptual and applied approaches designed to maintain alignment between the digital model and the real system over time. The analysis covers studies that introduce continuous validity monitoring mechanisms as well as works whose methodologies can be adapted for recurrent use. The review then examines how effectively these approaches address key challenges, such as variable selection, deviation detection, and identification of discrepancy sources.

The relevance of this research lies in providing an updated overview of emerging continuous or adaptive validation solutions, which remain incipient in the literature. By synthesizing these contributions and their limitations, the study highlights conceptual requirements and guidelines for the future development of more robust frameworks. Such advances are expected to enhance the reliability, traceability, and adaptability of DTs in dynamic industrial environments.

The remainder of this paper is organized as follows: Section 2 presents the theoretical background; Section 3 details the methodology; Sections 4.1 to 4.3 report the categorized review results; Section 5 provides a comparative discussion of the findings; and Section 6 presents the conclusions.

2. THEORETICAL BACKGROUND

2.1 Digital Twin validation and challenges of continuous operation

The validation of simulation models and DTs has traditionally been performed during the development phase, ensuring that the model adequately represents the real system under initial conditions. However, this initial validation does not guarantee that the DT remains faithful as the system evolves. Operational changes, the introduction of new operating patterns, or sensor degradation may gradually introduce discrepancies between the DT and the physical system. As a result, several authors emphasize the importance of adopting a lifecycle perspective for V&V in DTs.

Meng et al. [11] argue that model validity in continuous operation should be assessed periodically or continuously to ensure reliability. Zare and Lazarova-Molnar [12] reinforce this view by studying DTs in manufacturing environments with high human involvement, concluding that continuous validation requires accessible solutions capable of incorporating human expertise into the process.

Despite this recognition, the literature reports few initiatives involving active revalidation. In their systematic review, Bitencourt et al. [8] found that structured validation procedures are rarely implemented during the operational phase of DTs. Santos et al. [4] reached a similar conclusion in a bibliographic survey: the vast majority of DT projects neglect post-deployment validation, with only rare exceptions. This situation is partly explained by the lack of standardized methodologies and by the additional effort required for continuous validation. Conventional validation methods, such as comparisons based on aggregated metrics or sporadic statistical tests, face significant limitations in the context of operational DTs.

One major challenge concerns increasing dimensionality. When multiple system variables are monitored simultaneously, pointwise hypothesis tests become inefficient and prone to inflated Type I or Type II error rates. In addition, assessments based solely on descriptive statistics (i.e., comparisons of means or variances) may lead to misleading conclusions if data distributions or inter-variable correlations are ignored [13]. Another limitation of traditional approaches lies in neglecting performance history: treating each validation as an isolated event prevents the identification of trends or gradual drift in model behaviour.

In summary, pointwise and non-adaptive methods tend to fail in capturing subtle or progressive changes in the fidelity of a DT relative to the real system. These challenges highlight the need for continuous, automatable, and statistically grounded approaches for DT monitoring.

3. METHODOLOGY

This review follows a systematic procedure for literature search and analysis based on Mesquita et al. [14]. First, studies explicitly addressing continuous validation, recurrent monitoring, or adaptive updating of digital models and DTs were identified. Databases such as IEEE Xplore, Scopus, and Web of Science were queried using combinations of English keywords, including "digital twin", "discrete event simulation", "validation", and terms related to continuous monitoring approaches, such as "online validation", "continuous validation", and "real-time validation".

This search identified representative studies that consolidate the state of the art in DT verification and validation, including the systematic reviews by Bitencourt et al. [8], who analysed 250 papers; Santos et al. [4], with 75 papers; and Biazen et al. [9], with 72 papers. The applied study by Lugaresi et al. [13], which evaluated 12 publications focused on online validation of manufacturing models, was also considered. In this study, the term online refers to recurrent model updates performed in real time or at regular intervals, characterizing a continuous synchronization process between the physical and digital domains.

Next, inclusion criteria were applied to refine the scope: (i) studies proposing a method (conceptual or applied) related to model validation during the operational phase (rather than only initial validation); (ii) practical case studies or experiments illustrating continuous or adaptive validity monitoring; or (iii) review papers discussing challenges and gaps associated with lifecycle validation. Based on these criteria, 25 core studies were selected for in-depth analysis. Among them, approximately ten papers directly met criterion (i), that is, they proposed specific approaches for continuous monitoring of DT or simulation model validity. The remaining studies were included because they provide conceptual or technical foundations with potential for adaptation – criteria (ii) and (iii).

After study selection, the identified approaches were categorized using a strategy similar to that adopted in related works. Three main categories emerged from the analysis:

- Proposed continuous monitoring approaches, encompassing methods explicitly designed and published to continuously validate or monitor DTs (e.g., control chart–based methods, real-time similarity metrics, sequence alignment techniques). These studies are discussed in detail in the Development section of this paper.
- Methods with potential for recurrent use, including validation or verification techniques previously applied in a pointwise manner but that exhibit characteristics suitable for adaptation to continuous schemes, such as certain statistical procedures, signal analysis methods, and modular frameworks.
- Conceptually aligned approaches, covering frameworks and concepts (e.g., architectures and symbiotic systems) that, while not providing an operational method, support the relevance of and outline requirements for continuous DT validation.

This classification was iteratively refined during the critical reading of the selected papers. For each study, information was extracted regarding: (a) the methods or metrics employed; (b) application frequency (real-time continuous, periodic batch-based, or initialization-only); (c) application context (e.g., manufacturing, healthcare, logistics); and (d) main limitations or challenges reported. These aspects were then compared across studies, forming the basis for the synthesis presented in the Results Analysis section.

4. DEVELOPMENT

4.1 Existing approaches to continuous validation

In recent years, studies have begun to propose explicit methods for continuous validation or recurrent monitoring of DTs. This review identified works that introduce some form of active tracking of digital model validity during operation. Among them, approaches based on similarity metrics, event sequence analysis, and statistical control techniques stand out. The main proposals in this category are summarized below, with emphasis on their operating principles and applicability.

In broad terms, the contributions can be grouped into five main lines. The first involves the comparison of temporal sequences and traces, aiming to capture divergences in the dynamic behaviour of the system. Lugaresi et al. [13] propose online validation based on similarity between real and simulated events using Longest Common Sub-sequence (LCSS), Modified Longest Common Sub-sequence (mLCSS), and Dynamic Time Warping (DTW), enabling adherence monitoring even in the presence of noise and temporal misalignment. The second line focuses on co-evolution and adaptive updating approaches, in which the DT is continuously confronted with real data and adjusted when discrepancies exceed predefined thresholds. In this direction, Mertens and Denil [15] present a DT–physical system co-evolution workflow with dynamic triggers for parameter updates during operation.

The third line integrates machine learning and Statistical Process Control (SPC) to monitor validity over time. Santos et al. [16] combine a K-NN classifier with a *p*-type control chart to track DT accuracy, structuring a change-detection mechanism that links supervised patterns with statistical rules. Complementarily, Friederich and Lazarova-Molnar [17] propose a framework for data-driven discrete-event simulation models, defining revalidation policies (continuous, periodic, and conditional) and indicating SPC as a possible criterion to trigger revalidation, although without directly implementing control charts.

A fourth line addresses continuous validation through time- and usage-dependent credibility metrics, emphasizing the dynamic assessment of DT reliability. Lu et al. [18] introduce the Application-Time-Window (ATW) based metric, weighted by model usage frequency, while Lu et al. [19] extend this proposal by incorporating multiple time windows and a predictive agent to iteratively update DT credibility in an operational context.

Finally, some proposals place stronger emphasis on statistical and architectural aspects. Nikula et al. [20] propose continuous multivariate residual monitoring using the Mahalanobis distance with dynamic normalization adaptation, targeting autonomous deviation detection in metallurgical DTs. He et al. [21] adopt multivariate validation by epochs, organizing data according to operational regimes and applying statistical tests to verify DT behavioural stability under different conditions. In parallel, Zare and Lazarova-Molnar [22] discuss a modular strategy for validation and validity maintenance at the submodel level (with a focus on systems with AGVs), enabling selective policies for revalidation, recalibration, or re-extraction depending on the type of detected change. Mertens and Denil [23], in turn, advocate the reuse of classical V&V methods integrated into operational cycles, outlining guidelines for modular continuous validation in cyber-physical production systems.

Taken together, these studies consolidate the understanding that DT validation must go beyond pointwise practices and incorporate recurrent mechanisms, whether through temporal comparison, update routines, credibility metrics, or multivariate statistical monitoring. At the same time, the proposals vary in terms of statistical formalization and the degree of operational deployment in industrial settings, indicating that continuous validation remains in a phase of methodological consolidation.

To further clarify how these approaches can be operationalized in practice, an illustrative synthesis based on the reviewed literature can be considered. In a typical manufacturing scenario, a DT is used to monitor the performance of a production system, such as a machining cell or automated assembly line. During operation, real-time data collected from the physical system are continuously compared with simulation outputs generated by the DT.

Sequence-based techniques, such as LCSS or DTW, can be used to assess alignment between real and simulated event traces, while statistical monitoring methods, such as control charts or multivariate residual analysis, enable the detection of deviations over time. When discrepancies exceed predefined thresholds, adaptive updating mechanisms are triggered, allowing the DT to recalibrate its parameters or structure.

In parallel, time-dependent credibility metrics, such as those based on application time windows, can be used to assess the reliability of the DT throughout its lifecycle, considering both the timing and frequency of its use. This integrated workflow illustrates how the different validation strategies identified in the literature can be combined to support continuous monitoring and maintenance of DT validity in real operational environments.

4.2 Methods with potential for recurrent validation

Beyond works explicitly conceived for continuous validation, the literature also reports methods applied in a pointwise manner that could be adapted for recurrent use or periodic validation. This category includes techniques originally developed to validate models at specific points in time but whose processes are repeatable and therefore suitable for monitoring validity persistence over time. Approaches based on signal analysis and performance metrics are particularly noteworthy, as they could support continuous DT adherence assessment if automated.

These contributions can be organized into four methodological lines. The first encompasses frequency-domain approaches that compare spectral patterns of real and simulated signals. Lugaresi et al. [24] propose validating digital models by comparing power spectral densities (PSD) of performance indicators, enabling the identification of discrepancies in KPI trends and cycles. Morgan and Barton [25] extend this approach by employing the Fourier transform and metrics based on weighted mean coefficients, providing a quantitative criterion to detect changes in signal behaviour. Although these studies do not explicitly target online validation, both are compatible with continuous-flow applications, especially when runtime time series are available.

The second line comprises sequence and trace alignment methods aimed at structured comparisons of system behaviour over time. Muñoz et al. [26] use an adaptation of the Needleman–Wunsch algorithm to align state trajectories between the DT and the physical system, quantifying execution correspondence. From a more reactive-systems perspective, Leroy et al. [27] propose operators and trace comparison metrics (including editable distances such as Levenshtein), along with graphical representations that facilitate the identification of cycles and anomalies in state machines. In both cases, the main contribution lies in providing replicable comparative instruments that can be executed periodically or automatically within verification routines.

The third line corresponds to methods based on functional similarity metrics, applicable in mechatronic and robotics contexts. Gong et al. [28] develop motor similarity metrics to assess, in real time, fidelity between simulated motions and physical executions in human–robot interaction. Although the application is specific, the logic of functional validation through similarity metrics can be transferred to industrial scenarios in which the DT represents movements, trajectories, or operation patterns with a strong temporal component.

The fourth line gathers contributions focused on indicator-based monitoring and performance governance, emphasizing stability and DT usefulness over time. Overbeck et al. [29] use simple statistical indicators (means, deviations, and descriptive measures) to continuously compare performance between the physical system and the DT in manufacturing. Despite limitations in capturing complex multivariate variations, the approach stands out for its ease of implementation and alignment with industrial practice. Walton et al. [2], in turn, introduce metrics such as relevancy decay and effectiveness lag to assess the loss of effectiveness of DT-supported prescriptive decisions in dynamic environments. Although these metrics do not constitute a formal revalidation mechanism, they provide measurable bases for monitoring model "obsolescence" in real use. Complementarily, Casas [30] proposes a continuous VV&A process for simulation models, structured as iterative cycles of review and updating. While not necessarily operating in real time, the cyclical and systematic nature of this process offers a clear operational reference for adaptation to recurrent validity maintenance routines.

In summary, the studies in this group contribute by providing comparative metrics and verifiable routines that can be embedded into recurrent validation processes, even when originally conceived for pointwise use. Collectively, these approaches reinforce that the feasibility of continuous monitoring does not depend solely on explicitly "online" frameworks, but also on robust comparative methods capable of supporting decisions related to revalidation, recalibration, or updating throughout the DT operational lifecycle.

4.3 Conceptually aligned approaches to continuous validation

This section considers studies that, although not proposing a direct solution for continuous monitoring, are conceptually aligned with the idea of maintaining model validity in an adaptive manner. It includes frameworks and architectures that emphasize the need for recurrent DT updates, as well as studies that apply statistical techniques similar to those suggested for continuous validation, even if in different contexts. These works provide theoretical grounding and evidence of feasibility for lifecycle-oriented validation strategies.

A significant portion of these studies consistently shows that DT and simulation model validation remains predominantly pointwise and offline. Biazen et al. [9], in their review of simulation model verification and validation methods across multiple domains, highlight the scarcity of techniques designed to operate in a continuous or adaptive manner. A similar conclusion is reported by Bitencourt et al. [8], who analysed more than 250 studies on DT V&V in manufacturing and found that most approaches focus on the initial development phase, neglecting validity maintenance during operation. Both studies reinforce the need to integrate statistical methods, machine learning, and recurrent monitoring routines.

Other contributions emphasize validity as a dynamic property intrinsically linked to the system lifecycle. Boschert and Rosen [31] argue that DT fidelity should not be treated as a static attribute but rather as one that degrades as the physical system evolves. Complementarily, Kritzinger et al. [32], in their review of integration levels between physical and digital systems in I4.0, explicitly recognize that validation must be maintained dynamically, particularly in scenarios involving bidirectional integration and real-time use.

Organizational, human, and governance aspects also recur in this group. Zare and Lazarova-Molnar [12] discuss DT validation challenges in labour-intensive environments, where human variability limits the application of fully automated techniques. The authors advocate continuous and adaptive approaches that combine quantitative metrics with expert knowledge.

From a broader governance perspective, Schleich et al. [33] propose tracking data fidelity throughout the product lifecycle, emphasizing that model validity should evolve alongside the real system, even though they do not present a specific operational method.

Finally, conceptual frameworks seek to structure validation across the lifecycle. Waters [34] presents the TEVV (Testing, Evaluation, Verification, and Validation) framework for cyber-physical systems and DTs, with an emphasis on online validation and the detection of model degradation without interrupting operation. Although conceptual, this proposal reinforces the need for standards, ontologies, and systematic processes to sustain DT reliability under continuous use.

In summary, the works in this group converge in recognizing that the absence of systematic continuous validation methods constitutes a central gap in the literature. By explicitly highlighting limitations of existing approaches and framing validity as a lifecycle process, these contributions provide the conceptual basis for developing statistically grounded, automatable, and replicable methods for continuous DT validity monitoring.

5. RESULTS ANALYSIS

The comparative analysis presented in this section does not aim to propose an alternative solution, but rather to consolidate evidence of recurring limitations in existing approaches and to make explicit the open challenges associated with continuous DT validation.

As outlined earlier, this review was structured around the identified method categories. The subsection on Existing Approaches to Continuous Validation detailed proposals already established in the literature, covering techniques based on machine learning, statistical indicators, and event sequence alignment, among others. Methods with Potential for Recurrent Validation discussed techniques that, although originally applied in static validation settings, can be transformed into continuous procedures, including spectral signal analysis and modular model validation. Finally, Conceptually Aligned Approaches explored ideas and frameworks that reinforce the need for adaptive validation, such as adaptive DT architectures.

Across these subsections, the approaches vary in terms of maturity, some have been applied in real cases, while others remain largely conceptual, and in statistical scope, ranging from ad hoc methods to more general frameworks. Overall, the review covered studies from multiple sectors, primarily smart manufacturing, but also software engineering and generic cyber-physical systems. This diversity shows that concern with continuous model validity spans different domains, even though the available solutions remain fragmented.

Based on the reviewed approaches, recurring patterns and critical gaps in the state of the art of continuous DT validation become evident. The following analysis compares the identified methods and discusses their strengths and weaknesses relative to requirements highlighted in the literature.

1. Nature of Metrics and Deviation Detection: Existing methods differ in the type of metric used to assess validity. Some rely on global error or similarity metrics (e.g., classifier accuracy in Santos et al. [16], spectral distance in Lugaresi et al. [24]), while others evaluate pointwise differences between DT outputs and the real system (e.g., control charts in Overbeck et al. [29], event alignment in Lugaresi et al. [13]). Global metrics tend to simplify interpretation by providing a single fidelity indicator, but they may mask which specific aspects of behaviour are diverging. Pointwise difference approaches offer greater diagnostic granularity but require the definition of statistical thresholds or alarm criteria for multiple variables simultaneously. A recurring difficulty concerns the definition of these criteria: for example, Overbeck et al. [29] used ad hoc thresholds for mean differences, whereas Santos et al. [16] adopted classical control limits ($\pm 3\sigma$) in a *p*-chart. The lack of statistical standardization implies that each study empirically calibrates its deviation detector, hindering direct performance comparison across

methods. Overall, SPC-inspired methods offer a more formal framework for deviation detection, with explicit control of false-alarm probabilities, representing a conceptual advantage over purely ad hoc approaches.

2. Diagnostic Capability: A consistent gap lies in diagnosing the source of discrepancies once detected. Several authors acknowledge that their methods signal when “something is wrong” between the DT and the real system but do not indicate what or where the problem is. This limitation is explicitly noted in Santos et al. [16], who could not identify the variable responsible for alarms, and in Lugaresi et al. [13], whose review observed that most existing methods only flag anomalies without providing information about their origin. From a practical standpoint, diagnosis is crucial: once a system indicates that a DT has lost validity, engineers need to know which part of the model requires adjustment. Future methods should incorporate diagnostic mechanisms, whether through statistical decomposition, sensitivity analysis, or interpretability techniques (e.g., assigning SHAP weights to variables in ML models integrated with DTs). Without such mechanisms, users may hesitate to adopt continuous validation schemes due to concerns over “black box” alarms that do not guide corrective actions.

3. Model Adaptability and Updating: Detecting validity loss is only part of the challenge; the next step involves adapting or recalibrating the DT to restore accuracy. In this respect, most studies focus on monitoring itself and provide limited discussion on how models should be updated after an alarm. Some exceptions deserve mention: Lugaresi et al. [13] note that, upon detecting sequence divergence, the DT could be automatically resynchronized or relearned from observed events, although they do not detail the procedure. Ogunsakin et al. [35], in the context of Adaptive DTs, assume that adaptive components reoptimize the DT. In general, however, a clear link between detection and action is missing. Continuous validation approaches could draw inspiration from adaptive control systems by establishing triggers to initiate recalibration (e.g., re-estimating simulation model parameters) or by activating online learning modes that allow the DT to gradually adjust to new data. In practical terms, the absence of a feedback loop may limit the impact of proposed techniques: they signal validity loss but do not ensure validity maintenance unless an external agent intervenes. To evolve from reactive monitoring to truly adaptive validation, future research should integrate automatic DT update routines, ensuring a continuous cycle of detection, adaptation, and validation.

4. Complexity and Industrial Feasibility: A pragmatic aspect concerns implementation cost and feasibility in real industrial contexts. Methods requiring complex additional instrumentation, massive data collection, or high real-time processing capacity may face practical barriers. For example, the trace alignment technique proposed by Muñoz et al. [26] can be computationally intensive when applied to systems with thousands of daily events. Conversely, the method by Overbeck et al. [29] is simple and computationally inexpensive but may require frequent human intervention to analyse indicators and adjust limits. SPC-based approaches offer the advantage of being lightweight and well known in industry, facilitating adoption. Robustness to noise and outliers is also critical: real environments exhibit natural variability and noisy data, and overly sensitive validation systems may generate frequent false alarms, whereas overly permissive ones may miss subtle changes. None of the reviewed studies explicitly reported sensitivity analyses to noise or false-alarm rates. Therefore, there is room for future comparative studies evaluating common performance metrics (e.g., mean time to detection, false-alarm rate) across continuous methods to quantify their effectiveness. To date, the literature lacks standardized benchmarks, as each author tested methods on different cases, making it difficult to conclude which technique performs best in general. In this regard, initiatives such as Lugaresi et al. [13], which evaluated a dozen online validation methods, represent valuable steps toward building a comparative foundation.

5. Identified Gaps and Opportunities: Compiling the observations above highlights key gaps that must be addressed to advance continuous DT validation:

- Formal criteria for variable selection: Many studies monitored variables chosen empirically or for convenience. Systematic methods for defining which DT variables should be monitored to maximize effectiveness are lacking. Dimensionality-reduction techniques such as PCA could help identify variable combinations that are more sensitive to deviations. None of the reviewed works explicitly applied PCA or correlation analysis prior to monitoring.
- Multivariate versus univariate integration: Univariate approaches may miss deviations that only become evident through joint variable relationships. Few methods truly address multivariate monitoring. An opportunity lies in combining multivariate charts for global detection with univariate charts for visual support and diagnosis, leveraging both robustness and interpretability.
- Standardization and unified frameworks: Current solutions remain isolated. In this context, a valuable contribution would be the development of a unified framework incorporating multiple layers of continuous validation, such as outlier-aware data preprocessing, monitoring of critical variables, automatic model recalibration via incremental learning, and diagnostic reporting. Such a framework could build on existing quality engineering standards adapted to DT contexts. Bitencourt et al. [8] already identify fragmentation and lack of standardization as a key issue.
- Comprehensive case studies: Finally, large-scale implementations across different domains remain scarce. Most studies rely on simulated cases or controlled laboratory environments. Testing these approaches in real industrial settings, such as continuously operating factories, energy plants, or complex logistics systems, would be essential to validate practical applicability and uncover unanticipated challenges (e.g., data latency or unforeseen sensor changes). This transition to practice is necessary for continuous DT validation to move beyond research and become part of standard DT management toolkits.

6. CONCLUSION

The main contribution of this work is to demonstrate, based on a critical and systematic literature review, that continuous validation of DTs still lacks systematic methods, formal statistical criteria, and consolidated approaches for diagnosing deviations throughout the lifecycle. The analysis shows that viable approaches exist for continuously monitoring the validity of simulation-based DTs. In general, these approaches rely on comparisons between real and simulated data and on statistical metrics, but they remain fragmented and heterogeneous in terms of formalization and applicability.

However, current approaches still lack standardization and diagnostic robustness. Most do not provide formal mechanisms to identify which model component caused a detected divergence, nor do they establish clear procedures for acting upon an alarm (e.g., model recalibration or decision-maker notification). This limitation is critical for industrial adoption, as alert systems that do not provide actionable insights tend to be ignored over time.

In this context, continuous DT validation emerges as a strategic and growing field to ensure that the benefits promised by these technologies are sustained over time. Although still incipient, the reviewed literature already includes relevant contributions that demonstrate the technical feasibility of continuous DT validity monitoring. These studies indicate that it is possible to distinguish transient deviations from structural changes in system behaviour, reinforcing the potential of such approaches to support reliable DT operation in real environments.

The review also shows that isolated univariate approaches may be insufficient in complex scenarios with multiple correlated variables. Recent trends suggest combining multivariate methods for global anomaly detection with univariate methods for variable-level analysis. When coupled with contribution decomposition techniques, this integration appears promising for meeting simultaneous detection and diagnosis demands, enabling both the identification of anomalies and the localization of their sources within the DT.

From a conceptual standpoint, the literature converges on the view that DT validity is a dynamic attribute that must be continuously monitored and managed, alongside other system aspects such as performance and safety. Adaptive DT architectures and trust frameworks reinforce that validation should not be treated as a one-time event, but as an ongoing lifecycle process. This paradigm shift requires corresponding changes in the methodologies adopted by both industry and researchers.

Based on the gaps highlighted by this review, several future directions can be outlined. First, the literature points to the need for integrated frameworks that combine multiple layers of continuous validation, ranging from statistically guided variable selection and monitoring to feedback mechanisms that enable semi-automatic model updating when validity loss is detected. Second, the analysed studies indicate opportunities for adopting online machine learning approaches in combination with classical statistical techniques, allowing DTs to learn from deviations, for example, by recalibrating simulation model parameters as real data accumulate, thus maintaining accuracy without constant human intervention.

Another important avenue involves developing standardized performance metrics for continuous validation methods (such as detection time, detectable deviation range, and false-alarm rate) and conducting independent comparative studies. This would enable quantitative assessment of method suitability for specific systems or operational requirements. The creation of benchmark simulation environments, where different techniques can be tested under identical conditions, would further advance the field and increase industrial stakeholder confidence.

In conclusion, maintaining continuously validated DTs is challenging but increasingly necessary as DTs are deployed in critical and long-duration operations. The reviewed works provide a valuable starting point, showing that statistical monitoring of DT–physical system alignment and detection of significant deviations are feasible. To become part of standard practice, however, these solutions must evolve toward more comprehensive, reliable, and implementable frameworks. It is expected that convergent research efforts along these lines will yield consolidated methodologies that allow DTs to adapt to changes without losing fidelity, thereby enhancing reliability and traceability in I4.0 applications.

ACKNOWLEDGEMENTS

The authors acknowledge the support of UNIFEI, CAPES, CNPq and FAPEMIG.

REFERENCES

[1] Pereira, A. B. M.; Montevechi, J. A. B.; Pinto, W. G. M.; Santos, C. H. (2023). Simulation and digital twins to support reverse logistics decisions: a review, *International Journal of Simulation Modelling*, Vol. 22, No. 3, 381-391, doi:10.2507/IJSIMM22-3-640

[2] Walton, R. B.; Ciarallo, F. W.; Champagne, L. E. (2024). A unified digital twin approach incorporating virtual, physical, and prescriptive analytical components to support adaptive real-time decision-making, *Computers & Industrial Engineering*, Vol. 193, Paper 110241, 14 pages, doi:10.1016/j.cie.2024.110241

[3] Wright, L.; Davidson, S. (2020). How to tell the difference between a model and a digital twin, *Advanced Modeling and Simulation in Engineering Sciences*, Vol. 7, No. 1, Paper 13, 13 pages, doi:10.1186/s40323-020-00147-4

[4] Santos, C. H. dos; Montevechi, J. A. B.; Queiroz, J. A. de; Miranda, R. de C.; Leal, F. (2022). Decision support in productive processes through DES and ABS in the Digital Twin era: a systematic literature review, *International Journal of Production Research*, Vol. 60, No. 8, 2662-2681, doi:10.1080/00207543.2021.1898691

[5] Harper, A.; Mustafee, N. (2019). A hybrid modelling approach using forecasting and real-time simulation to prevent emergency department overcrowding, *Proceedings of the 2019 Winter Simulation Conference*, 1208-1219, doi:10.1109/WSC40007.2019.9004862

[6] Sargent, R. G. (2013). Verification and validation of simulation models, *Journal of Simulation*, Vol. 7, No. 1, 12-24, doi:10.1057/jos.2012.20

[7] Zhuang, C.; Liu, J.; Xiong, H. (2018). Digital twin-based smart production management and control framework for the complex product assembly shop-floor, *The International Journal of Advanced Manufacturing Technology*, Vol. 96, Nos. 1-4, 1149-1163, doi:10.1007/s00170-018-1617-6

[8] Bitencourt, J.; Wooley, A.; Harris, G. (2025). Verification and validation of digital twins: a systematic literature review for manufacturing applications, *International Journal of Production Research*, Vol. 63, No. 1, 342-370, doi:10.1080/00207543.2024.2357741

[9] Biazen, M. A.; Woldeyohannes, A. D.; Gebeyehu, S. G. (2025). Simulation models verification and validation: recent development and challenges: a review, *International Journal of Modeling, Simulation, and Scientific Computing*, Vol. 16, No. 1, Paper 2530001, 24 pages, doi:10.1142/S1793962325300018

[10] Hua, E. Y.; Lazarova-Molnar, S.; Francis, D. P. (2022). Validation of digital twins: challenges and opportunities, *Proceedings of the 2022 Winter Simulation Conference*, 2900-2911, doi:10.1109/WSC57314.2022.10015420

[11] Meng, C.; Nageshwaraniyer, S. S.; Maghsoudi, A.; Son, Y.-J.; Dessureault, S. (2013). Data-driven modeling and simulation framework for material handling systems in coal mines, *Computers & Industrial Engineering*, Vol. 64, No. 3, 766-779, doi:10.1016/j.cie.2012.12.017

[12] Zare, A.; Lazarova-Molnar, S. (2024). Validation of digital twins in labor-intensive manufacturing: significance and challenges, *Procedia Computer Science*, Vol. 238, 623-630, doi:10.1016/j.procs.2024.06.070

[13] Lugaresi, G.; Gangemi, S.; Gazzoni, G.; Matta, A. (2023). Online validation of digital twins for manufacturing systems, *Computers in Industry*, Vol. 150, Paper 103942, 14 pages, doi:10.1016/j.compind.2023.103942

[14] Mesquita, R. P.; Leal, F.; Queiroz, J. A. (2024). Digital twins in the retail industry: a systematic literature review, *International Journal of Simulation Modelling*, Vol. 23, No. 3, 424-434, doi:10.2507/IJSIMM23-3-690

[15] Mertens, J.; Denil, J. (2023). Digital-twin co-evolution using continuous validation, *Proceedings of the ACM/IEEE 14th International Conference on Cyber-Physical Systems (ICCPS '23)*, 266-267, doi:10.1145/3576841.3589628

[16] Santos, C. H. dos; Campos, A. T.; Montevechi, J. A. B.; Miranda, R. de C.; Costa, A. F. B. (2024). Digital twin simulation models: a validation method based on machine learning and control charts, *International Journal of Production Research*, Vol. 62, No. 7, 2398-2414, doi:10.1080/00207543.2023.2217299

[17] Friederich, J.; Lazarova-Molnar, S. (2023). A framework for validating data-driven discrete-event simulation models of cyber-physical production systems, *Proceedings of the 2023 Winter Simulation Conference*, 2860-2871, doi:10.1109/WSC60868.2023.10407382

[18] Lu, H.; Zhang, L.; Wang, K.; Cheng, H. (2024). An application-time-window based credibility assessment method for equipment digital twins in manufacturing, *Journal of Industrial Information Integration*, Vol. 41, Paper 100635, 10 pages, doi:10.1016/j.jii.2024.100635

[19] Lu, H.; Zhang, L.; Deen, M. J.; Wang, K.; Cheng, H.; Yang, L. T. (2025). Online credibility assessment of equipment digital twin for discrete manufacturing, *IEEE Transactions on Automation Science and Engineering*. Vol. 22, 14763-14774, doi:10.1109/TASE.2025.3563246

[20] Nikula, R.-P.; Remes, A.; Kaartinen, J.; Kortelainen, J.; Loponen, T.; Ruuska, J.; Ruusunen, M. (2025). Autonomous residual monitoring of metallurgical digital twins, *Minerals Engineering*, Vol. 220, Paper 109107, 13 pages, doi:10.1016/j.mineng.2024.109107

[21] He, L.; Rhodes-Leader, L.; Song, E. (2024). Digital twin validation with multi-epoch, multivariate output data, *Proceedings of the 2024 Winter Simulation Conference*, 347-358, doi:10.1109/WSC63780.2024.10838742

[22] Zare, A.; Lazarova-Molnar, S. (2024). Modular validation within digital twins: a case study in reliability analysis of manufacturing systems, *Proceedings of the 2024 Winter Simulation Conference*, 2903-2914, doi:10.1109/WSC63780.2024.10838952

[23] Mertens, J.; Denil, J. (2025). Reusing model validation methods for the continuous validation of digital twins of cyber-physical systems, *Software and Systems Modeling,* Vol. 24, No. 5, 1427-1449, doi:10.1007/s10270-024-01225-6

[24] Lugaresi, G.; Aglio, G.; Folgheraiter, F.; Matta, A. (2019). Real-time validation of digital models for manufacturing systems: a novel signal-processing-based approach, *Proceedings of the IEEE 15th International Conference on Automation Science and Engineering*, 450-455, doi:10.1109/COASE.2019.8843082

[25] Morgan, L. E.; Barton, R. R. (2022). Fourier trajectory analysis for system discrimination, *European Journal of Operational Research*, Vol. 296, No. 1, 203-217, doi:10.1016/j.ejor.2021.05.052

[26] Muñoz, P.; Wimmer, M.; Troya, J.; Vallecillo, A. (2024). Measuring the fidelity of a physical and a digital twin using trace alignments, *IEEE Transactions on Software Engineering*, Vol. 50, No. 12, 3122-3145, doi:10.1109/TSE.2024.3462978

[27] Leroy, D.; Bousse, E.; Megna, A.; Combemale, B.; Wimmer, M. (2018). Trace comprehension operators for executable DSLs, *Proceedings of the 14th European Conference on Modelling Foundations and Applications* (ECMFA 2018), 293-310, doi:10.1007/978-3-319-92997-2_19

[28] Gong, L.; Chen, B.; Xu, W.; Liu, C.; Li, X.; Zhao, Z.; Zhao, L. (2022). Motion similarity evaluation between human and a tri-co robot during real-time imitation with a trajectory dynamic time warping model, *Sensors*, Vol. 22, No. 5, Paper 1968, 14 pages, doi:10.3390/s22051968

[29] Overbeck, L.; Le Louarn, A.; Brützel, O.; Stricker, N.; Lanza, G. (2021). Continuous validation and updating for high accuracy of digital twins of production systems, *Proceedings of the ASIM-Fachtagung Simulation in Produktion und Logistik*, 609-617

[30] Casas, P. F. (2023). A continuous process for validation, verification, and accreditation of simulation models, *Mathematics*, Vol. 11, No. 4, Paper 845, 25 pages, doi:10.3390/math11040845

[31] Boschert, S.; Rosen, R. (2016). Digital twin – the simulation aspect, Hehenberger, P.; Bradley, D. (Eds.), *Mechatronic Futures*, Springer, Cham, 59-74, doi:10.1007/978-3-319-32156-1_5

[32] Kritzinger, W.; Karner, M.; Traar, G.; Henjes, J.; Sihn, W. (2018). Digital twin in manufacturing: a categorical literature review and classification, *IFAC-PapersOnLine*, Vol. 51, No. 11, 1016-1022, doi:10.1016/j.ifacol.2018.08.474

[33] Schleich, B.; Anwer, N., Mathieu, L., Wartzack, S. (2017). Shaping the digital twin for design and production engineering, *CIRP Annals*, Vol. 66, No. 1, 141-144, doi:10.1016/j.cirp.2017.04.040

[34] Waters, G. (2025). Testing, evaluation, verification and validation (TEVV) of digital twins: a comprehensive framework, *arXiv* preprint, doi:10.48550/arXiv.2507.04555

[35] Ogunsakin, R.; Mehandjiev, N.; Marin, C. A. (2023). Towards adaptive digital twins architecture, *Computers in Industry*, Vol. 149, Paper 103920, 18 pages, doi:10.1016/j.compind.2023.103920