Response to the Reviewer's Evaluation

Title of PhD thesis: Reliable Digital Twin Operation via Surrogate-Based and Online Learning-Driven Flowsheet Model Maintenance

Reviewer: Dr. Gyula Ádám Dörgő; Data Intelligence Department Manager; Borouge International

Dear Dr. Gyula Ádám Dörgő,

Thank you for your thorough and constructive evaluation of my PhD dissertation, as well as your positive assessment of its methodological contributions and scientific results.

My responses are provided in blue text directly below the corresponding sections of your review.

Sincerely,

Balázs Palotai

1. General evaluation:

Balázs Palotai's outstanding work offers significant novelty in the increasingly relevant topic of reliable and automated digital twin applicatons in the process industry. The work systematically identifies crucial pain points in the practical application of digital twins (primarily presented in the chapter „Moving towards digital twins in the process industry") and presents corresponding solutions to three of these pain points. Each solution is presented in a separate chapter: the chapter "Surrogate-based calibration of flowsheet models" formulates flowsheet model maintenance as an adaptive calibration problem and introduces a surrogatebased methodology that significantly reduces the computational cost, the chapter "Online learning for incremental surrogate model maintenance" examines how different online update mechanisms can sustain surrogate accuracy under evolving conditions, while the chapter "Integrated framework for adaptive flowsheet model maintenance" integrates the previously presented surrogate-based flowsheet calibration and online surrogate learning into a coherent framework. Each of these chapters concludes with case studies: a heat exchanger network model provides continuity and comparability between the different chapters, while two industrial case studies are presented in Chapter 5. In addition to the industrial case studies, a software prototype presented in Chapter 6 further underpins the intention and readiness of practical applications. The systematically and incrementally improving nature of these chapters perfectly demonstrates how the toolboxes of process engineering, data analysis and machine learning form a powerful synergy which is universally applicable to enhance the chemical industry, moreover, they highlight the widespread knowledge of the candidate in the fields of chemical engineering and data-based modelling.

It was particularly insightful to observe how the candidate identified complex practical problems in the modeling of chemical technologies and the application of these models as digital twins, elevated these problems to a higher level of integration, and proposed practical solutions after data-based evaluation. The chosen methods are well justified, the case studies and results are thoroughly evaluated, and the conclusions drawn are sound and clearly demonstrate the candidate's professional competence. Overall, the work was a pleasure to read.

I would like to express my sincere gratitude for your thorough evaluation of my PhD dissertation and your highly positive assessment of its scientific contributions, methodological approach, and practical relevance.

I particularly appreciate your recognition of the systematic development of the proposed methodologies and the integration of chemical engineering, process modeling, and machine learning. Establishing a connection between these disciplines and translating the resulting methodological developments into practically applicable solutions has been a central objective of my PhD research.

2. Formal and structural evaluation:

From a structural perspective, after the introduction, a more structural summary of the digital twins is presented, which is followed by the three methodological improvement proposals providing the main findings of the thesis work. The software prototype is presented in a separate chapter projecting towards the further industrial applications. As all of these chapters describe independent and standalone analysis results, which in addition build on top of each other, in my view, the selection of this structure is favorable and advantageous.

The sections build on the already accepted journal articles of the candidate, underpinning the scientific foundations of the results. The main part of the thesis consists of 152 pages and contains 149 references. It is well-edited and constructed in a transparent and easy-to-follow structure. The figures are high-quality and well-explained in the text.

Thank you for your positive evaluation of the dissertation's structure, organization, and presentation. I would also like to express my gratitude for the constructive comments and valuable suggestions provided during the previous evaluation round and discussed during the internal defense. These discussions helped me identify opportunities to further improve the clarity, consistency, and scientific presentation of the dissertation. I carefully considered the earlier remarks and incorporated the corresponding revisions into the final version, which I believe have significantly strengthened the overall quality and coherence of the work.

3. Evaluation of the thesis statements:

All theses are well-formulated, transparent, and clearly state the novelty and candidate's contribution. All of them are supported by at least one internationally accepted research paper, and some have already received independent citations, showing the cascading effect of the research.

Thesis #1: "I developed a framework supporting the industrial adoption of digital twins and demonstrated that one of the most critical prerequisites for their successful deployment and long-term operation is the automated maintenance of the underlying flowsheet models." (3 sub-theses)

The thesis results were published in a journal article and five conference abstracts.

Thesis #2: "I demonstrated that the proposed surrogate-based calibration methodology substantially reduces the computational cost of flowsheet model calibration while improving its robustness without compromising the required accuracy, thereby supporting the efficient application of flowsheet models in digital twin systems." (3 sub-theses)

The results of the thesis were published in a journal article and one conference full paper, for which the candidate has received an independent citation according to the MTMT.

Thesis #3: "I demonstrated that online (incremental) learning enables the continuous adaptation of surrogate models to changing operating conditions while preserving the previously acquired global knowledge of the system. Furthermore, I demonstrated that the timing of model updates and the selection of training data are independent design decisions, allowing different update strategies to be freely combined." (3 sub-theses)

The results of the thesis were published in a journal article and two conference abstracts.

Thesis #4: "I developed an integrated framework that combines surrogate-based flowsheet model calibration with online learning of surrogate models, thereby reducing the computational effort required for flowsheet model calibration and improving calibration efficiency, while effectively supporting the long-term adaptation of flowsheet models in digital twin systems." (3 sub-theses)

The results of the thesis were published in a journal article, one conference full paper and two conference abstracts, for which the candidate has received an independent citation according to the MTMT.

I sincerely appreciate your positive evaluation of the four thesis statements. I would also like to thank you for your constructive comments and suggestions regarding the thesis statements during the previous evaluation round. Your feedback was particularly valuable in refining their formulation and more clearly distinguishing the scientific novelty and individual contributions.

4. Questions:

Chapter 1 states that "Despite their conceptual maturity and demonstrated potential, DTs are significantly less prevalent in the process industry than in discrete manufacturing", along with the key issues of the application of digital twins. If, in theory, we had infinite resources to tackle the problem, what would be the candidate's roadmap to sustainably solve it? And what would be the first element to address: the technological, computational, organizational, financial, or any other aspects?

Assuming that unlimited resources also reflect a strong management commitment and a clear strategic vision for digital twin adoption, I would prioritize the organizational aspects as the first element of the roadmap.

The initial step would be to establish a dedicated, multidisciplinary organization bringing together process engineers, simulation specialists, data engineers, data scientists, and business representatives, while maintaining close collaboration with IT departments. This organization would take ownership of the development, deployment, and long-term operation of digital twin solutions, ensuring continuous engagement with end users. Its primary responsibility would be to translate operational and business needs into practical applications and ensure that the insights generated by digital twins are translated into informed decisions and concrete operational actions, rather than remaining unused results on dashboards.

I consider this organizational foundation essential for establishing trust and credibility among stakeholders. A clearly defined organizational structure, supported by management commitment, transparent responsibilities, and accessible expertise, would demonstrate that digital twins are not isolated research or digitalization projects, but reliable tools integrated into everyday operational and decision-making processes. In my view, this trust is a fundamental prerequisite for widespread user adoption.

Once this organizational foundation is established, the next priority would be to develop a comprehensive governance and life cycle management framework. The multidisciplinary organization would be responsible for defining common standards, managing model and data dependencies, establishing verification and validation procedures, and ensuring continuous monitoring of both model performance and realized business benefits.

Based on this foundation, technological and computational challenges could be addressed systematically through standardized data integration, modular and reusable model architectures, appropriate multi-fidelity modeling approaches, and automated model maintenance methodologies, including the surrogate-based calibration and online learning methods investigated in this dissertation.

Finally, the roadmap would focus on gradually extending digital twin applications across processes and organizational boundaries, supported by continuous user engagement, knowledge sharing, and evaluation of operational benefits. Even with unlimited resources, I would favor an incremental implementation strategy, as it allows the organization to validate

solutions, learn from practical experience, and progressively build confidence in the technology.

Therefore, my first priority would be organizational readiness and ownership, followed by governance and life cycle management, technological integration, and scalable implementation. I believe that technological and computational limitations can largely be addressed through appropriate expertise and investment, whereas sustainable adoption ultimately depends on establishing an organizational environment in which digital twins are continuously maintained, trusted, and actively used to support operational decisions and generate measurable business value.



Budapest, 2026.11.04

Balázs Palotai