Thesis points of doctoral (PhD) dissertation

Reliable Digital Twin Operation via Surrogate-Based and Online Learning-Driven Flowsheet Model Maintenance

Balázs Palotai

University of Pannonia

Chemical Engineering and Material Sciences Doctoral School

Supervisors:

Dr. Ágnes Takács-Bárkányi

Dr. Gábor Kis

Consultant:

Dr. Tibor Chován



Department of Process Engineering

Veszprém

2026

1. Background and objectives

Growing economic, energy, and sustainability requirements in the process industry have increased the demand for model-based decision support systems. However, existing solutions often rely on isolated models that are difficult to integrate, limiting information sharing and model reuse. The digital twin concept addresses these limitations by providing a unified framework that integrates models, data sources, and applications while allowing the virtual model to continuously evolve with the real process.

In the process industry, flowsheet models provide a suitable foundation for digital twins because they are based on first principles, are validated, interpretable, and trusted by engineers. However, their predictive accuracy gradually deteriorates due to changing operating conditions, equipment aging, and process drift, making continuous model maintenance essential. The lack of automated maintenance remains one of the main barriers to the industrial adoption of flowsheet-based digital twins.

The objective of this research was to develop a methodological framework for the long-term automated maintenance of flowsheet models, thereby supporting the reliable operation of digital twins. To achieve this, I identified the main barriers to industrial digital twin deployment and developed an integrated framework combining surrogate-based calibration with online learning for adaptive flowsheet model maintenance.

2. Thesis points

I. 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.

I demonstrated that the limited industrial adoption of digital twins in the process industry is primarily not the consequence of a single technological limitation, but rather of the combined challenges of model integration, life cycle management, and model maintenance.

Based on these findings, I developed a life cycle-oriented framework supporting the industrial deployment of digital twins. The framework defines the development stages from a digital model to a fully integrated digital twin, together with the associated technological, organizational, and operational requirements. Furthermore, it identifies the roles of DevOps, DataOps, MLOps, and model engineering practices in supporting model integration, reusability, and life cycle management.

I further demonstrated that, in the process industry, flowsheet models provide a suitable foundation for digital twin implementation due to their physical basis, interpretability, and engineering credibility. At the same time, I showed that one of the most important prerequisites for the long-term reliable operation of digital twins is the automated maintenance of flowsheet models, since the gradual degradation of model accuracy directly compromises the reliability and practical applicability of the digital twin.

Related publications: 1, 7, 8, 9, 10, 11

II. 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.

I demonstrated that the predictive accuracy of flowsheet models gradually deteriorates during operation due to process changes, such as equipment aging, heat exchanger fouling, catalyst deactivation, and variations in feedstock composition, making periodic recalibration essential. I further showed that the high computational cost and limited robustness of conventional direct flowsheet-based calibration significantly hinder the automated maintenance of flowsheet-based digital twins.

I developed a surrogate-based calibration methodology in which the computationally intensive flowsheet model is replaced by a neural network surrogate model during the optimization process. The methodology was further enhanced by a composite objective function that combines engineering priorities with measurement uncertainty, thereby improving calibration robustness.

The case study demonstrated that the proposed methodology reduced the computational cost of calibration by 96% while preserving the required accuracy and maintaining robust performance under noisy measurement conditions. These results confirm that surrogate-based calibration provides an effective solution for the automated maintenance of flowsheet models, thereby establishing one of the key prerequisites for the long-term reliable operation of digital twins.

Related publications: 2, 5

III. 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.

I demonstrated that the predictive accuracy of offline-trained surrogate models may deteriorate when the real process deviates from the operating region covered by the original training dataset. Since the objective of the offline training dataset is to construct a globally accurate surrogate model over the entire expected operating range, the resulting model does not necessarily provide sufficient accuracy in local operating regions that become dominant during long-term industrial operation.

I developed an online learning framework that enables the incremental updating of surrogate models using newly acquired operational data. I demonstrated that the controlled replay of samples from the original offline training dataset allows the surrogate model to adapt to new operating conditions while preserving previously acquired global knowledge, thereby preventing catastrophic forgetting.

Furthermore, I demonstrated that the timing of model updates and the selection of data used for model updating represent two independent design dimensions. Consequently, different online learning strategies can be flexibly combined according to application-specific requirements. The case study confirmed that the proposed methodology continuously improved local prediction accuracy while preserving the global predictive performance of the surrogate model over long-term operation.

Related publications: 3, 12, 13

IV. 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.

I demonstrated that the independent application of surrogate-based calibration and online learning is insufficient to ensure the long-term adaptation of flowsheet models. Although surrogate models substantially reduce the computational cost of calibration, their predictive accuracy gradually deteriorates when the real process deviates from the original offline training domain.

I developed an integrated closed-loop framework in which surrogate-based flowsheet model calibration and online learning mutually reinforce each other. Each successful flowsheet model calibration provides new training information for the surrogate model, which continuously adapts to changing operating conditions through online learning while preserving its previously acquired global knowledge.

The case study demonstrated that the proposed integrated methodology reduced the number of direct flowsheet model calibrations by 92% while maintaining the long-term predictive accuracy of the surrogate model. As a result, the surrogate model is transformed from a static approximation into a continuously maintained model that co-evolves with the flowsheet model, thereby effectively supporting the long-term adaptation of flowsheet models and the reliable operation of digital twins.

Related publications: 4, 6, 14, 15

3. Potential applications and future research

The results of this research can be directly applied to industrial digital twin systems that combine flowsheet-based process models with machine learning techniques for real-time decision support. The integration of surrogate-based calibration and online learning enables the rapid and reliable updating of first-principles models while significantly reducing computational cost. This is particularly beneficial for process units where equipment aging, heat exchanger fouling, catalyst deactivation, feedstock variability, or other nonlinear phenomena require frequent model recalibration. Owing to its flexibility, the proposed methodology can be readily extended to heat exchanger networks, distillation systems, reactors, and other process units where continuous model maintenance is essential.

Within digital twin-based decision support systems, the proposed framework enables more reliable process optimization, reduced operating costs, and condition-based operation of industrial equipment. The automated, data-driven model maintenance process allows the digital twin to continuously adapt to changing process, economic, and operating conditions without requiring manual model intervention. Furthermore, continuously updated surrogate models provide a foundation for predictive functionalities such as early fault detection, maintenance scheduling, and rapid evaluation of alternative operating scenarios.

The proposed methodology also offers broader opportunities beyond model maintenance. A continuously updated surrogate model can

serve as a computationally efficient approximation for optimization, operating window exploration, multi-objective economic optimization, and rapid what-if analysis. Within this framework, the flowsheet model acts as a first-principles reference for validating surrogate model predictions, thereby ensuring physical consistency and engineering credibility. This synergistic combination of first-principles and data-driven models provides a foundation for faster and more reliable decision support and offers potential for future applications in process intensification, investment planning, and predictive maintenance.

List of publications

Publications which form the basis of the doctoral dissertation

Publications in peer-reviewed journals:

1. B. Palotai, Á. Bárkányi, G. Kis, and J. Abonyi. Moving Towards Digital Twins: Overview of Key Challenges in the Process Industry. *Heliyon, 12 (14) (2026) e45350*.
   Scimago Journal Ranking: Q1, Impact factor: 3.6

2. B. Palotai, G. Kis, J. Abonyi, and Á. Bárkányi. Surrogate-based flowsheet model maintenance for digital twins. *Digital Chemical Engineering, 15 (2025) 100228.*
   Scimago Journal Ranking: Q1, Impact factor: 5.4

3. B. Palotai, G. Kis, T. Chován, and Á. Bárkányi. Online learning for surrogate model maintenance. *Hungarian Journal of Industry and Chemistry, 53 (2) (2025) 37–43.*
   Scimago Journal Ranking: Q4, Impact factor: 0.8

4. B. Palotai, G. Kis, T. Chován, and Á. Bárkányi. Online learning supported surrogate-based flowsheet model maintenance. *Digital Chemical Engineering, (2025) 100287*. Scimago Journal Ranking: Q1, Impact factor: 5.4

Full papers in international conference proceedings:

5. B. Palotai, Á. Bárkányi, G. Kis, and J. Abonyi. Operationalization management: Enhancing life cycle management of digital twins. *Computer Aided Chemical Engineering, 53 (2024) 3025–3030*.

6. B. Palotai, G. Kis, J. Abonyi, and Á. Bárkányi. Synergies of adaptive learning for surrogate-based flowsheet model maintenance. *ESCAPE 35 – European Symposium on Computer Aided Process Engineering, Ghent, Belgium, 6–9 July 2025. ISBN: 978-9-492859-36-5*.

International conference:

7. B. Palotai, G. Kis, Z. Mathe. Unlocking value through operationalization and advanced digital twin applications in downstream. *Emerson User Exchange EMEA 2024, Düsseldorf, Germany, 2024*.

8. B. Palotai, G. Kis, Z. Mathe. Digital twins in action: Joint effort of Hysys and DeltaV to make profit real. *Emerson Exchange 2025, San Antonio, Texas,USA, 19–22 May 2025*.

9. B. Palotai, G. Kis, L. Grad-Gyenge, Z. Máthé, T. Chován, and Á. Bárkányi. Toward transparent decision support: Building fully integrated digital twin for refinery operations. *Advances 2025, Manchester, UK, 16–17 October 2025*.

10. B. Palotai, G. Kis, L. Grad-Gyenge, Z. Máthé, T. Chován, and Á. Bárkányi. Building a fully integrated digital twin for refinery operations. *IChemE Process Management and Control Special Interest Group Webinar, Online, 27 March 2026.*

National conference:

11. B. Palotai. Moving towards digital twins in the process industry. *30th Process Control Systems Meeting (PCS), Miskolc-Lillafüred, Hungary, 14–16 October2024*.

12. B. Palotai, G. Kis, T. Chován, and Á. Bárkányi. Online learning for surrogate model maintenance. *Műszaki Kémiai Napok'25, Veszprém, Hungary, 15–17April 2025.*

13. B. Palotai, G. Kis, T. Chován, and Á. Bárkányi. Exploring AI and Machine Learning-Based Decision Support Systems for Real Chemical Processes. *University Research Scholarship Program (EKÖP) Institutional Conference, University of Pannonia, Veszprém, Hungary, June 25, 2025*.

14. B. Palotai, G. Kis, T. Chován, and Á. Bárkányi. Maintenance of flowsheet models in digital twins through real-time data-driven learning in the chemical industry. *Műszaki Kémiai Napok'26, Veszprém, Hungary, 21–23 April 2026*.

15. B. Palotai, G. Kis, T. Chován, and Á. Bárkányi. Exploring AI and Machine Learning-Based Decision Support Systems for Real Chemical Processes. *University Research Scholarship Program (EKÖP) Institutional Conference, University of Pannonia, Veszprém, Hungary, June 25, 2026*.