Operationalization Management: Enhancing Life Cycle Management of Digital Twins

Balázs Palotai,ab\* Ágnes Bárkányi,b Gábor Kis,a János Abonyib

aMOL Group Plc., Dombóvári Street 28, Budapest, H-1117, Hungary

bUniversity of Pannonia, Egyetem Street 10, Veszprém, H-8200, Hungary \*bpalotai@mol.hu

Abstract

The recent progress in development of Information Technology (IT) gave rise to a new wave of industrial transformation marked by cloud computing, the Industrial Internet of Things (IIoT), Big Data analytics, Industry 4.0 principles, and autonomous systems. Digital Twins are at the core of this revolution, by bridging physical world with its digital representation to optimize Cyber-Physical Production Systems (CPPS) in order to create more value. However, it is quite challenging to validate that of the anyway obvious theoretical advantages even in the case of a pilot project not to mention a full production unit size Digital Twin. Another aspect of challenges is the need for model life-cycle management emerges to preserve the benefit captured by the new Digital Twin based technologies. This paper introduces a novel methodology inspired by Operations-based frameworks and Model Engineering, addressing these bottlenecks. It offers a unique solution for managing simulation models monitoring and maintenance in Digital Twins applications. The paper shows the benefit of surrogate-based automated flowsheet model fitting solution for a simplified refinery case-study to reduce the expensive simulation use for model fitting, and reduced the time required compared with the direct simulation fitting without losing accuracy.

**Keywords**: Digital Twin, Life cycle, SimOps, Surrogate

* 1. Introduction

Industrial processes, encompassing from chemical manufacturing to power generation, have long relied on simulation and optimization to enhance efficiency, cost-effectiveness, and safety. Traditional manual process simulation models, built with precise attention to detail, have been fundamental in achieving these goals. These models, based on the governing physical and thermodynamic principles, have provided critical insights, enabling informed decision-making. This widespread use of flowsheet process simulation models made them perfect constituents of Digital Twin solutions which naturally brings the need for standardization and automation of model creation, maintenance, and utilization. The emergence of Machine Learning Operations (MLOps) offers a promising solution for this purpose as showed by D. Kreuzberger et al. Initially designed for machine learning models, MLOps is now expanding to encompass first-principal process simulation models (SimOps) as demonstrated by the work of I. Pan et al.

In this paper, we present an evolution of process simulation, transitioning from labor-intensive manual methods to MLOps-driven automation and efficiency. We highlight the most critical challenges in traditional simulation model maintenance and propose an automated solution. This solution combines the precision of first-principal models with the agility of modern data-driven and surrogate approaches. In this paper we show the benefit of proposed surrogate-based automated flowsheet model fitting solution through a simplified refinery case-study to reduce the expensive simulation use for model fitting.

* 1. General introduction of Digital Twin life cycle management

Digital Twins represent an innovative paradigm where digital counterparts of physical systems bridge the gap between production processes, enabling dynamic data collection for real-time monitoring, asset health analysis, performance evaluation, and informed decision-making as it is presented by the work of Singh et al. This seamless synchronization between digital and physical systems, whether in an online or offline context, is the key to continuous optimization and monitoring within production systems.

In the refining sector, and in the broader process industry, flowsheet simulation historically takes a pivotal role. These tools are instrumental in the detailed examination of process technologies, the execution of "what if" scenario analyses, process optimization, and equipment health monitoring. Based on these capabilities they can provide an acceptable foundation for Digital Twin applications.

The literature highlights a critical challenge in Digital Twin (DT) adoption: a lack of trust among stakeholders, hindering effective implementation as it is described in the work of Müller et al. To resolve this problem, it's essential to select sufficiently simple and understandable project(s) for demonstration of DT benefits which can facilitate, building trust and acceptance from the future users. Once benefits are confirmed, the focus shifts to seamless DT integration with existing processes, supporting business decisions. Maintaining transparency and credibility is key for value creation (Figure 1). To achieve this, proper life cycle management is a prerequisite.

Given the intricate nature of DT frameworks, leveraging methodologies such as DevOps, MLOps, DataOps, and SimOps are instrumental for their life cycle management. DevOps offers a comprehensive framework for overseeing the entire DT life cycle, ensuring quality and continuous improvement. MLOps facilitates the development and performance management of machine learning models within DTs. DataOps, a process-driven methodology, provides data-focused analytics and sustains data quality throughout the life cycle. Similarly, SimOps, which has the same function as MLOps, supports the life cycle management of simulation models as proposed by I. Pan et al. MLOps can serve as a robust guideline to introduce SimOps for maintaining flowsheet process simulation models within the Digital Twin framework, as the model data requirements and building steps align closely with ML model development methodologies (Figure 2).



Figure 1. Digital Twin steps and AI readiness combined with model life cycle management



Figure 2. Simplified MLOps methodology with proposed SimOps connection points

The key steps in MLOps and SimOps for flowsheet simulation models (FSMs):

* **Data Pipeline**: Digital Twin frameworks employing steady-state flowsheet models require, additional steps become necessary like steady-state detection and data reconciliation as showed by B. Farsang et al.
* **Simulation Model Building**: Creating, fine-tuning, and evaluating simulation models often necessitates manual modeling expertise. While research into automated flowsheet identification is ongoing as showed by M. Barth et al., manual input most probably will remain essential.
* **Model Monitoring**: Various approaches exist for detecting concept/model drift during the operation as summarized by F. Bayram et al.
* **Model Maintenance**: As FSMs are primarily manually generated, therefore development and maintenance of industrial applications based on these models can be laborious and costly as highlighted by G. S. Martínez et al. While data-driven approaches can be maintained with less engineering effort by applying MLOps methodology.

From our literature review, it becomes evident that addressing flowsheet model drift is a pivotal component of FSM life cycle management. If concept drift arises from operational degradations like fouling in heat exchangers, separation efficiency reduction, or catalyst deactivation, the model maintenance process could be automated by adopting MLOps methodology. To address this, we introduce a novel solution for the automated re-training of flowsheet process simulation models.

* 1. Proposed automated model re-training solution

Model retraining practically an optimization challenge, focusing on minimizing the residual error between the model's estimated and observed parameters. However, using optimization algorithms with flowsheet models the complexity increases significantly. Some of these models can demand extensive CPU computational resources. Even when computational time is not excessive, accurately estimating derivatives for gradient-based algorithms becomes problematic due to the noise introduced by these black box models. This noise can originate from factors like small variable sensitivities, algorithm termination criteria or model stability.



Figure 3. Sour-water stripper unit Hysys model and the observed fouling effect

When practicality dictates that not practical to use flowsheet models for optimization, an alternative approach is to treat the original model as a source of "computational experiments," generating data points as if physical experiments had been conducted. These data are then used to construct simpler models that employ explicit functions, often referred to as surrogate, reduced-order, or metamodels. Many research efforts have utilized this surrogate approach for the optimization of flowsheet-based operations as it is showed by J. A. Caballero et al. Despite the advantages, so far just limited research has been carried out on the use of surrogate approaches for model fitting type of optimization.

To illustrate the advantages of using surrogate models for model fitting as a partial inverse model, we employed Aspen Hysys steady-state simulation for sour-water stripper unit model (Figure 3), wherein a frequently occurring heat exchanger fouling phenomenon was investigated. Historical data indicated a decrease in column inlet temperature and an increase in lean water temperature which caused by the increasing fouling in feed heat-exchanger. To simplify the problem and focus now solely on surrogate-based model fitting, we used the historical steady-state daily average operational data after model-based data reconciliation as a fitting dataset and tried to identify the fouling parameter for each day by minimizing the residual error between the validation and model estimated temperatures. The applied process workflow summarized on Figure 4.

* + 1. Data Generation and Surrogate Training

Utilizing the base Hysys model, we generated 500 simulation data points for training an Artificial Neural Network (ANN) surrogate within the operational envelope, using Latin Hypercube Sampling (LHS). The ANN was designed with input parameters, including boundary conditions (such as feed quality and mass-flow), manipulated parameters (e.g., reflux rate), and the fitting factors, such as the Fouling factor. The ANN output parameters included independent variables like column inlet temperature and lean water outlet temperature, steam consumption, etc. The trained surrogate model consistently achieved an impressive average R-squared value of 0.935 on test data (30%). For clarity and accessibility, we stored both the Hysys and surrogate models, as well as the associated training data, in a model registry and model datastore, respectively.

* + 1. Optimization Process

The optimization phase involved the utilization of the Particle Swarm Optimization (PSO) algorithm. The cost function, the Mean Square Error (MSE), quantified the difference between measured and predicted key parameters (in this scenario, column inlet and lean water temperature were used) and the acceptance limit was defined based on practical modelling considerations. The optimization started with the surrogate model (full operation envelope & 20 particles), which proposed optimal fitting parameters for the daily fitting data (90 data points). These proposed parameters were subsequently cross verified against the original Hysys model.



Figure 4. Surrogate-based model fitting workflow

If the simulation MSE fell below the predefined threshold, the outcome, along with state conditions, was stored in the fitting datastore. Conversely, if the threshold was not met, the optimization continued with employing the simulation model within a narrow parameter range (+/- 10% & 5 particles). If this secondary simulation result met the threshold, it was added to the fitting datastore, and the corresponding simulation data was incorporated into the surrogate training dataset, enhancing model accuracy. In cases where the threshold proved elusive, a warning raised the need for manual intervention to investigate the anomaly.

* + 1. Results and discussion

Leveraging the surrogate model's exceptional accuracy, the optimization process mostly proposed optimal fitting parameters (Figure 5). Additional simulation optimization was only required when the surrogate model's accuracy fell under the desired threshold. In summary, the surrogate model not just contributed to cost reduction by minimizing the need for expensive simulations in model fitting but also reduced the overall time required for the process (Figure 6), without compromising the accuracy of the results. Utilizing additional simulation-based optimization results, the overall accuracy of surrogate model can be further improved and consequently the expensive simulation-based optimization usage can be reduced. It's important to emphasize that the primary role of the surrogate model was not to estimate fouling factors but to replace the simulation model.



Figure 5. Fitting results by datapoints and fitting steps



Figure 6. Fitting time consumption below and above the threshold

This versatility allows the surrogate model to serve not only in model maintenance but also as a valuable asset in operational optimization and model-based data reconciliation. As model complexity escalates, the importance of adaptive model sampling and surrogate retraining becomes increasingly pronounced within the Digital Twin framework.

* 1. Conclusions

Based on the literature review, the flowsheet model drift handling is the most critical part of life cycle management of flowsheet models based Digital Twin framework. If the concept drift caused by operational degradations like fouling in heat-exchangers, reduced separation efficiency or catalyst deactivation, the model maintenance can be automated.

The introduced novel hybrid surrogate-based automated flowsheet model fitting solution yielded promising results for a simplified refinery case-study and demonstrated the benefits of using surrogate models in the Digital Twin framework. Applying this method, the use of expensive simulation can be significantly reduced in comparison to the time required by the direct simulation fitting without losing accuracy.

References

Barth, Mike, and Alexander Fay. "Automated generation of simulation models for control code tests." Control Engineering Practice 21.2 (2013): 218-230.

Bayram, Firas, Bestoun S. Ahmed, and Andreas Kassler. "From concept drift to model degradation: An overview on performance-aware drift detectors." Knowledge-Based Systems 245 (2022): 108632.

Caballero, José A., and Ignacio E. Grossmann. "An algorithm for the use of surrogate models in modular flowsheet optimization." AIChE journal 54.10 (2008): 2633-2650.

Farsang, Barbara, Sándor Németh, and János Abonyi. "Role of steady state data reconciliation in process model development." Hungarian Journal of Industry and Chemistry 41.1 (2013): 65-75.

I. Pan, L. R. Mason, O. K. Matar, Data-centric Engineering: Integrating simulation, machine learning and statistics. Challenges and opportunities, Chemical Engineering Science 249 (2022) 117271

Kreuzberger, Dominik, Niklas Kühl, and Sebastian Hirschl. "Machine learning operations (mlops): Overview, definition, and architecture." IEEE Access (2023).

Martínez, Gerardo Santillán, et al. "Automatic generation of a high-fidelity dynamic thermal-hydraulic process simulation model from a 3D plant model." IEEE Access 6 (2018): 45217-45232.

Müller, Julian Marius. "Antecedents to digital platform usage in Industry 4.0 by established manufacturers." Sustainability 11.4 (2019): 1121.

Singh, Maulshree, et al. "Applications of digital twin across industries: A review." Applied Sciences 12.11 (2022): 5727.