

Accuracy Comparison of Data-Driven Modeling Techniques for a Modified Quadruple Tank Process in Energy-Efficient Control

Ernadia Rosman¹, Sazuan Nazrah Mohd Azam^{1*}, Arfah Syahida Mohd Nor¹,
Muhammad Nizam Kamarudin¹, Zakariah Yusuf²

¹Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka, Melaka, Malaysia

²Faculty of Electrical Engineering, Universiti Teknologi MARA, Cawangan Johor, Pasir Gudang, Malaysia

Email: *sazuan@utem.my

How to cite this paper: Rosman, E., Azam, S.N.M., Nor, A.S.M., Kamarudin, M.N. and Yusuf, Z. (2025) Accuracy Comparison of Data-Driven Modeling Techniques for a Modified Quadruple Tank Process in Energy-Efficient Control. *Journal of Power and Energy Engineering*, 13, 243-260. <https://doi.org/10.4236/jpee.2025.139017>

Received: August 11, 2025

Accepted: September 21, 2025

Published: September 24, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In energy critical systems, accurate modeling of fluid dynamics is essential for designing controllers that enhance operational performance and reduce energy consumption. This study presents a data-driven modeling framework for the Modified Quadruple Tank Process (MQTP) utilizing three distinct system identification techniques: Transfer Function (TF) State-Space (SS) and Process Model (PM) to comparatively assess their accuracy in capturing the system's dynamic behavior. Each model is evaluated based on its ability to replicate the nonlinear interactions and predict liquid level variations across the four interconnected tanks. The TF model exhibits high fidelity for the lower tanks (T1 and T2) achieving Best-fit values of 88.37% and 84.88% respectively. While its accuracy is lower for the upper tanks (T3 and T4), with a Best-fit value of 67.77%, it remains sufficient for the intended control application, staying within acceptable error thresholds. Closed-loop validation using Mean Absolute Percentage Error (MAPE) and Normalized Root Mean Square Error (NRMSE) confirms the TF model's effectiveness in tracking liquid levels particularly in T1 and T2 (MAPE = 10.05% and 9.13%). Despite lower performance in T3 and T4, the model remains within acceptable error thresholds (MAPE ≤ 16%). The study contributes to the modeling of MQTP systems by demonstrating the viability of integrating transfer function models with real-time experimental data, addressing a critical gap in predictive modeling for complex fluid dynamics in energy applications. These findings support the development of energy-efficient control strategies through accurate system representation.

Keywords

Modified Quadruple Tank Process, Data-Driven Modeling, System

1. Introduction

Data driven modeling has become a powerful approach in engineering and system dynamics, especially when first-principles models are difficult to derive. Unlike traditional models, it uses historical and real-time data to capture complex behaviors without deep physical insights. This method is advantageous for nonlinear, stochastic, or high-dimensional systems. Its adaptability and scalability enable real-time prediction, control, and diagnostics in industries. For example, machine learning is used in predictive maintenance, fault detection, and control optimization, improving efficiency and reliability [1] [2].

Furthermore, the evolution of Industry 4.0 has pushed the demand for data-centric intelligence, integrating data-driven models into cyber-physical systems, smart factories, and digital twins [3]. In such settings, data-driven models provide a low-cost and flexible alternative to physics-based simulations, particularly where sensor networks generate massive amounts of real-time operational data. They also support hybrid modeling approaches where physical laws are partially known but complemented with statistical learning for improved accuracy and robustness [4]. This confluence of computational power, accessible data, and advanced algorithms has positioned data-driven modeling at the core of modern system engineering innovation.

The quadruple-tank system is widely recognized as a benchmark multivariable process control problem used to demonstrate advanced control strategies, particularly in scenarios involving multiple-input multiple-output (MIMO) configurations. A modified version of this system introduces additional nonlinearities and structural flexibility, making it even more representative of real-world industrial dynamics such as fluid level control, chemical mixing, and distributed energy systems. The key feature of the system lies in its controllable non-minimum phase behavior, which presents a challenge for conventional control schemes and thus serves as an ideal platform for testing robust and intelligent control algorithms [5].

The modified system enhances modeling complexity by enabling switchable configurations or introducing actuator constraints, delay characteristics, and varying flow interactions. These properties allow the system to emulate complex plants with interacting subsystems, particularly relevant for process industries and smart infrastructure environments. Additionally, the modular nature of the system supports experimental validation, making it a popular choice for laboratory testbeds in academic and industrial control research. For example, decentralized fault-tolerant fuzzy controllers (PI-FLC), sliding mode FOPID schemes, and nonlinear observer-based backstepping have been implemented for stability enhancement in such systems [6] [7].

In energy-related domains, the modified quadruple tank model is increasingly applied for validating fluid distribution strategies in thermal energy storage systems, where accurate liquid level control directly influences system efficiency and safety. Its dynamic and nonlinear nature ensures that any modeling or control approach developed on this platform demonstrates robustness, adaptability, and scalability—critical traits for deployment in modern control environments [8].

Energy systems, especially in the era of smart grids and decentralized resources, are increasingly characterized by their complexity, variability, and scale. In this context, data-driven modeling has emerged as a transformative approach for improving the operational efficiency and resilience of energy infrastructures. These models are essential in capturing dynamic behaviors across subsystems like generation, storage, and distribution, enabling real-time monitoring, predictive control, and intelligent decision making. Notably, they allow system operators to bypass incomplete or highly nonlinear physical models, replacing them with adaptive structures trained from historical or streaming data [9].

One of the most impactful applications of data-driven techniques lies in microgrids and distributed energy resource (DER) management. By leveraging predictive algorithms, operators can anticipate load demands, optimize storage dispatch, and adjust renewable integration strategies under uncertainty [10]. This results in reduced operational costs, enhanced energy reliability, and lower carbon footprints. Moreover, data-driven control has been instrumental in the predictive maintenance and stability assurance of complex power systems, ensuring seamless integration of fluctuating renewables such as wind and solar energy [11].

Practical applications include smart building energy management using regression-based model predictive control (MPC), reinforcement learning for HVAC, and surrogate modeling for district energy. These highlight the adaptability of data-driven models and their ability for detailed optimization across scales. The rise of sensors and IoT devices increases data flows, boosting the importance of machine learning and data analytics in energy systems. Studying data-driven modeling and system identification will influence energy-efficient control, emphasizing the need for precise modeling to develop controllers that maximize energy efficiency.

2. Methodology

The methodology is structured into four primary phases: Experimental Setup and Data Acquisition, Data Preprocessing, Model Development, and Model Validation and Performance Evaluation. These phases collectively form the foundation for building a data-driven dynamic model for the MQTP, crucial for understanding the system's behavior for energy-efficient control.

2.1. Experimental Setup and Data Acquisition

The experimental setup is designed to replicate real-world conditions for the MQTP, which consists of four interconnected tanks, two upper tanks (T3 and T4) and two lower tanks (T1 and T2). Liquid is circulated by two pumps (P1 and P2), where

P1 delivers liquid to T1 and T4, and P2 supplies T2 and T3. The system relies on gravity flow, resulting in significant cross-coupling effects for system modeling. **Figure 1** illustrates the tank arrangement of MQTPP.

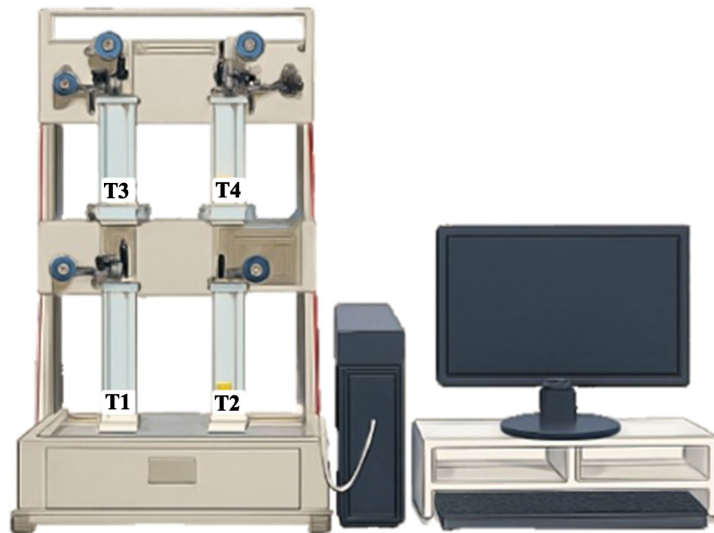


Figure 1. Detail labelling for tank arrangement in MQTPP.

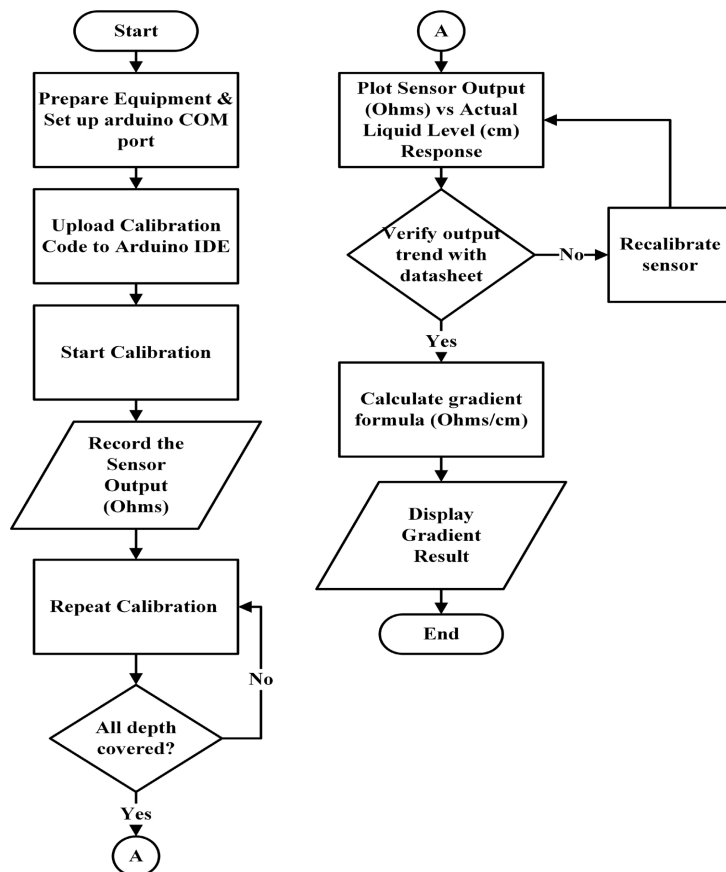


Figure 2. Flowchart of calibrating the sensors in MQTPP.

The liquid levels are measured with eTape sensors connected to an Arduino system for real-time data. The sensors are calibrated as per manufacturer guidelines for accuracy. The system records real-time input-output data, crucial for developing dynamic models, capturing the link between pump actions and liquid levels. Figure 2 illustrates the calibration procedure.

Figure 3 and Figure 4 display sensor output trends, comparing readings to liquid level changes. This graph illustrates the resistance-liquid level relationship, validating calibration accuracy and data reliability during testing.

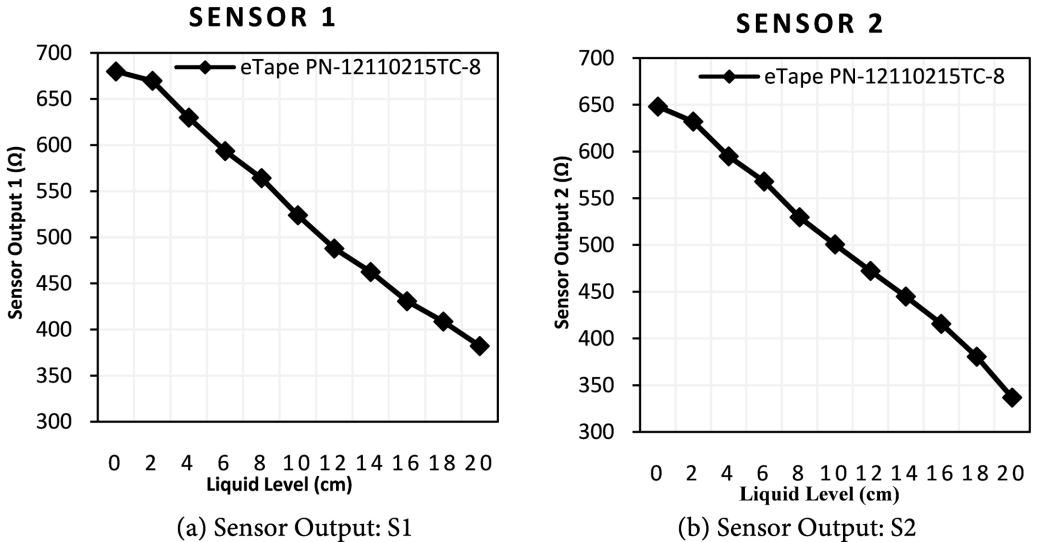


Figure 3. Sensor output trend for sensor in lower tanks.

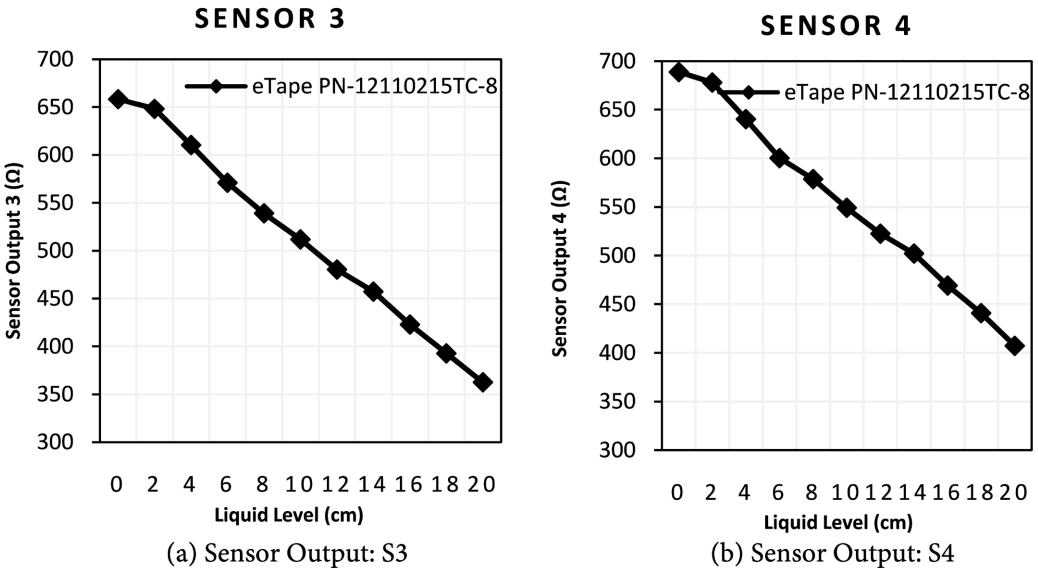


Figure 4. Sensor output trend for sensor in upper tanks.

Data collection uses a closed-loop system to keep liquid levels within desired limits. The feedback adjusts control inputs based on sensor data to prevent over-

flow and maintain stability. As shown in **Figure 5**, this setup captures real-world system dynamics.

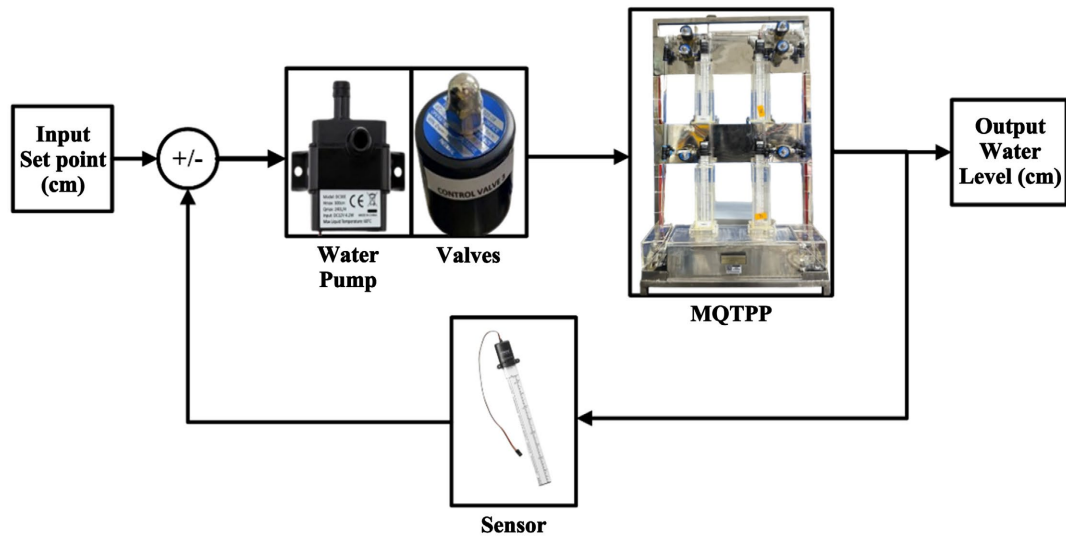


Figure 5. Configuration of the closed-loop system for data collection.

Two scenarios were tested using ten excitation multi-step inputs to analyze the impact of pumps P1 and P2 on specific tanks. By activating only one pump at a time and monitoring each tank’s dynamic response, input-output data were gathered for system identification. This data reveals the interactions between the pumps and tanks, which is essential for building the system model. The details of the two scenarios are illustrated in **Figure 6**.

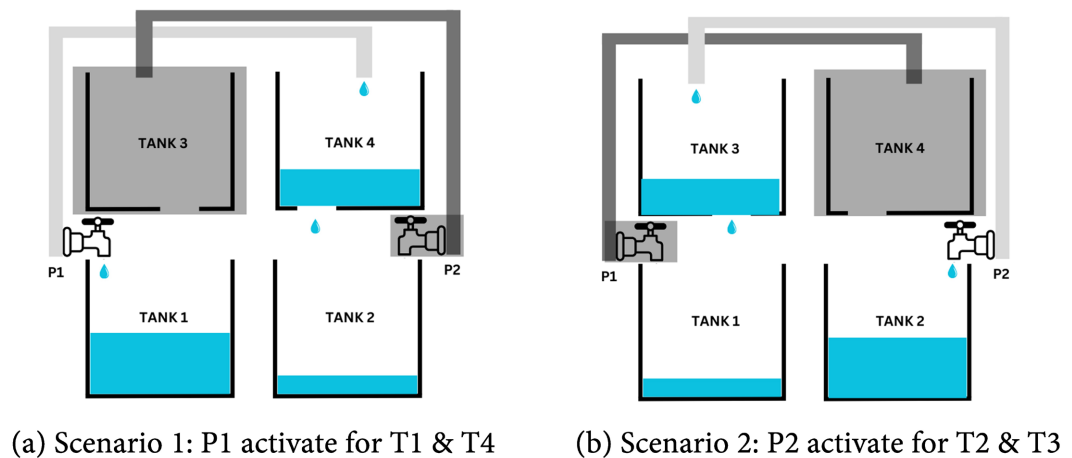


Figure 6. Operating scenario for closed-loop data collection.

During Scenario 1, P1 supplies liquid to T1 and T4, with T4’s outflow feeding T2. The goal is to study T1 and T4 behavior while isolating other tanks. The grey component indicates T3 and P2 are unaffected. Though T2 has an inflow, it’s negligible due to low flow and short duration. The main aim was to isolate T1 and

T4's response to P1. Initial observations showed T4's outflow minimally affects T2's level, not impacting system behavior. This simplification helps identify input-output relations without added complexity. In Scenario 2, P2 activates, affecting T2 and T3.

2.2. Data Preprocessing

Data preprocessing is critical in preparing the raw experimental data for system identification and model development. The raw data typically contains noise and fluctuations that can obscure the underlying system dynamics. Therefore, it is essential to perform preprocessing steps to ensure the data is clean, reliable, and suitable for further analysis.

The preprocessing steps are as follows:

1) Denoising: Using a Bayesian denoising approach, the first step involves removing high-frequency noise from the raw data. This approach minimizes false anomaly detections while preserving key trends and system behavior variations.

2) Filtering: A low-pass filter with a cutoff frequency of 0.50 Hz is applied to remove high-frequency components from the signal. This preserves the low-frequency data crucial for analyzing the long-term system dynamics.

3) Smoothing: The Savitzky-Golay method smooths the filtered data, reducing residual errors and minimizing fluctuations. This step ensures that the data reflect the actual underlying trends without over-smoothing, which could remove the significant variations [12] [13].

Table 1 outlines a detailed method for data preparation ahead of analysis. These preprocessing steps help ensure that the data used for system identification is accurate and free from noise, allowing for the development of reliable models.

Table 1. Parameter setting for all stages of processing raw data.

Step	Parameter settings	
	Parameter	Settings
Denoising	Method	Bayes
	Levels	11
	Threshold	Soft
	Noise estimate	Level dependent
Low pass filter	Frequency	0.50 Hz
Savitzky-Golay	Smoothing factor	0.05

2.3. Model Development

Model development is essential for capturing the dynamic behavior of the MQTP system. This process uses the real-time data collected during the experiments to develop a data-driven dynamic model. The MATLAB System Identification Toolbox is employed to identify models that represent the system's dynamics based on the input-output data.

Several model structures are evaluated to determine the most suitable representation for the system. Transfer Function (TF) models use polynomial ratios to describe the system's input-output relationship. These models are simple to interpret and provide insights into the system's dynamic behavior [14]. State-Space (SS): SS models are more comprehensive and capture the system's internal state. They use differential equations to describe the system's multivariable dynamics and are more suitable for complex systems [15]. Process Models (PM): PM models provide a simplified physical description of the system, including parameters such as gain, time constants, and delays. Although less complex, they are useful for real-time applications where computational efficiency is essential [16].

To guarantee unbiased and dependable model estimation, the full dataset of 6492 samples was split equally into two subsets [17]:

- 1) 50% of datasets used for estimation (1 - 3246 samples)
- 2) 50% of datasets used for validation (3247 - 6492 samples)

This data splitting strategy, shown in **Figure 7**, helps to prevent overfitting. Overfitting occurs when a model fits the training data perfectly but struggles to generalize to new or unseen data.

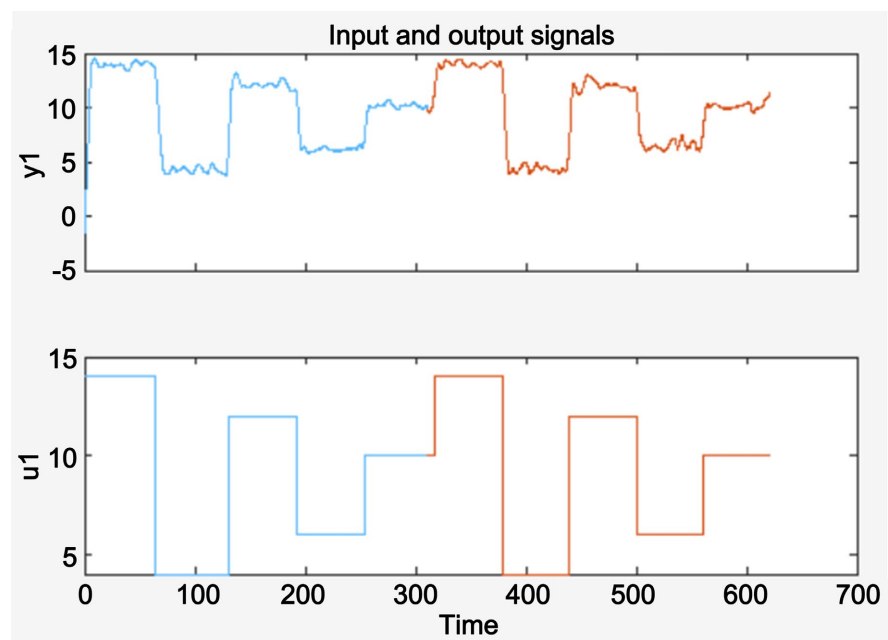


Figure 7. The splitting strategy in input-output signals.

2.4. Model Validation and Performance Evaluation

In this phase, the identified model from the Best-fit evaluation was then assessed by applying it in a closed-loop system to observe its performance compared to the real MQTP plant. The model was tested with the same single setpoint applied to the simulated TF model in Simulink and the MQTP real plant, directly comparing their behavior under identical conditions. **Figure 8** illustrated the Simulink block diagram of closed-loop simulation identified model, and the representation of

closed-loop system applied in MQTP is same as **Figure 5**.

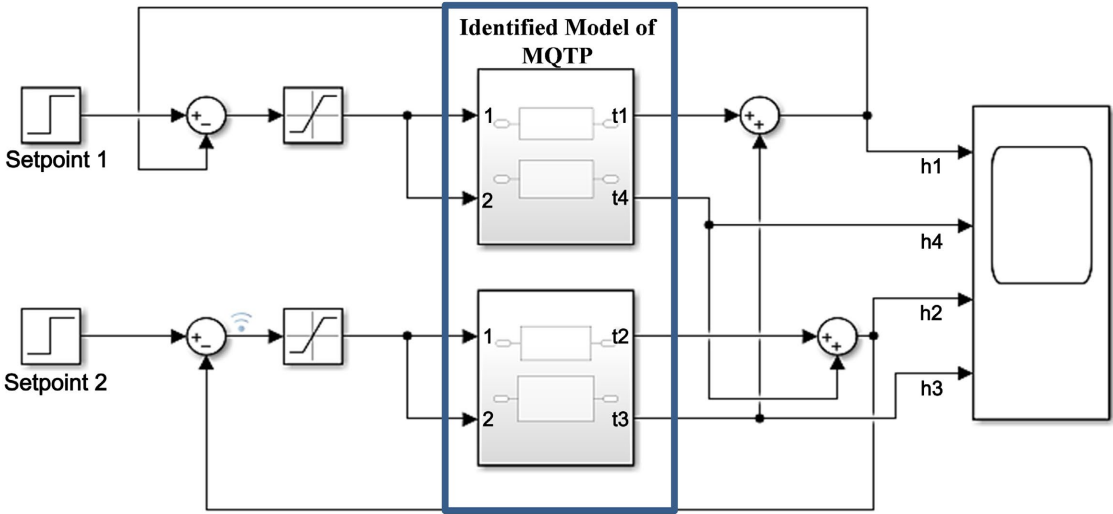


Figure 8. Closed-loop simulation in MATLAB.

Two common metrics for model validation are, Mean Absolute Percentage Error (MAPE): This metric calculates the average magnitude of errors between the predicted and actual outputs, expressed as a percentage of the actual data. It provides a clear measure of the model’s prediction accuracy [9].

$$MAPE(\%) = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_{sim(i)} - y_{real(i)}|}{y_{real(i)}} \right) \times 100 \tag{1}$$

where:

- y_{sim} : The predicted output from the Simulink model
- y_{real} : The actual output from the MQTPP
- n : The number of data points

Normalized Root Mean Squared Error (NRMSE): This metric quantifies the magnitude of errors by taking the square root of the average squared differences between predicted and observed values. The NRMSE is normalized to ensure that it can be compared across different datasets or models, making it a robust measure of model accuracy [18].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{sim(i)} - y_{real(i)})^2} \tag{2}$$

$$NRMSE = \frac{RMSE}{y_{max} - y_{min}} \tag{3}$$

where:

- y_{min} : The minimum observed values
- y_{max} : The maximum observed values

Figure 9 points out the conceptual framework for the validation process implemented for closed-loop testing for both simulation and real-plant.

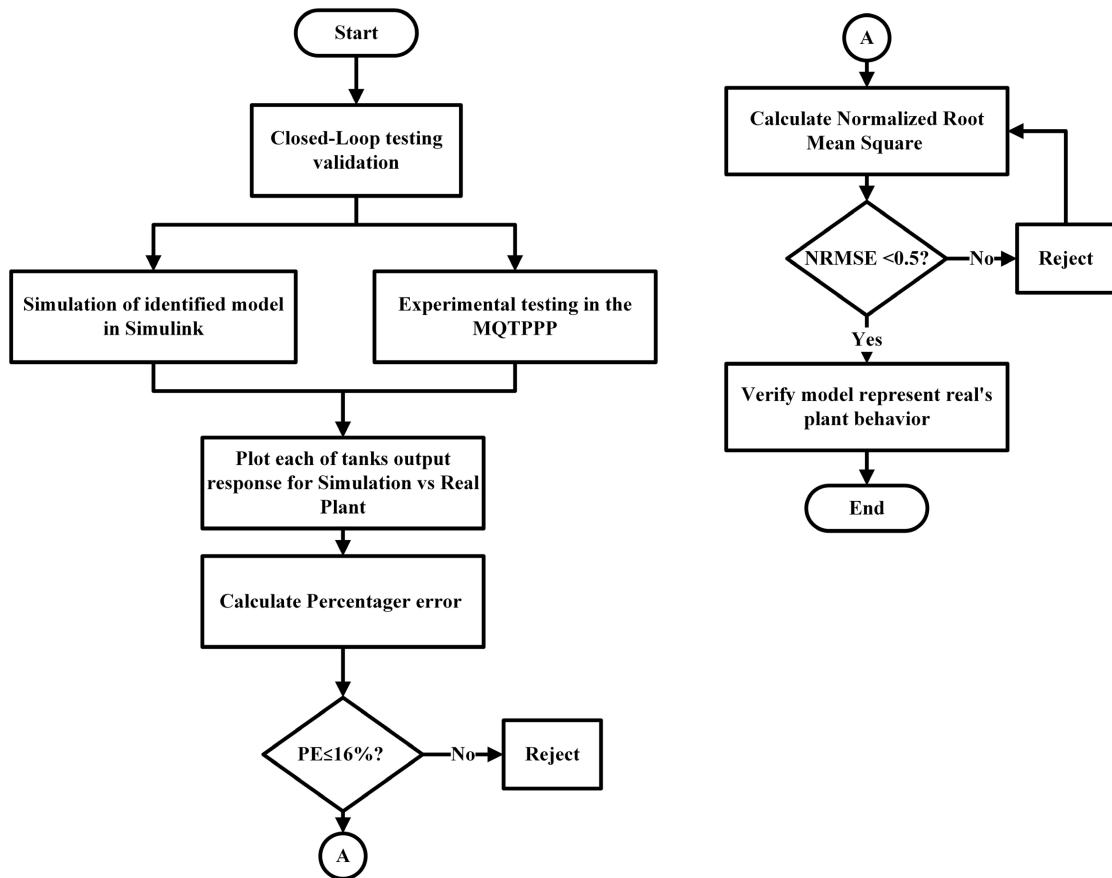


Figure 9. Flowchart for closed-loop validation upon simulation vs real plant.

3. Results and Discussion

This section presents the results of the system identification, model development, and validation processes. This discussion focuses on model performance evaluation based on Best-fit values, a comparative analysis of different modeling structures, closed-loop validation with experimental data, and the broader implications for energy applications.

3.1. Model Performance Evaluation

The performance of the identified models was evaluated using the Best-fit criterion, which measures the agreement between the predicted model output and the experimental data. The Best-fit values indicate how well the model represents the system's dynamics, with higher values corresponding to better model accuracy. These values were derived by splitting the data as being mentioned in subsection 2.3.

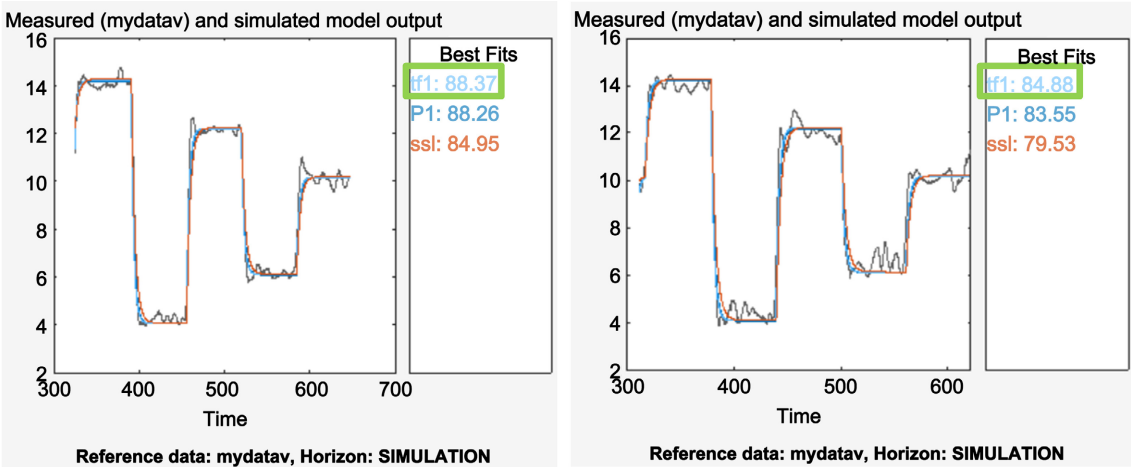
For the transfer function (TF) model, the Best-fit values were 88.37% for T1, 84.88% for T2, and 67.77% for T3 and T4. These results suggest the TF model performs well for the lower tanks (T1 and T2). Still, it struggles to capture the complexities of the upper tanks (T3 and T4), where the system dynamics involve stronger nonlinear interactions and cross-coupling effects.

In comparison, the State-Space (SS) model showed the Best-fit values of 84.95%

for T1, 79.53% for T2, and 29.44% for T3 and T4. While the SS model provided a good fit for the lower tanks, its performance for the upper tanks was significantly lower, suggesting that although the SS model accounts for system dynamics with internal states, it still struggles to capture the upper tanks' complexities fully.

The Process Model (PM) exhibited Best-fit values of 88.26% for T1, 83.55% for T2, and 40.63% for T3 and T4. This model performed similarly to the TF model for the lower tanks, but its accuracy decreased significantly for the upper tanks. The PM, being a simplified model, provides computational efficiency but lacks the complexity required to capture the detailed dynamics of the system, particularly in the upper tanks.

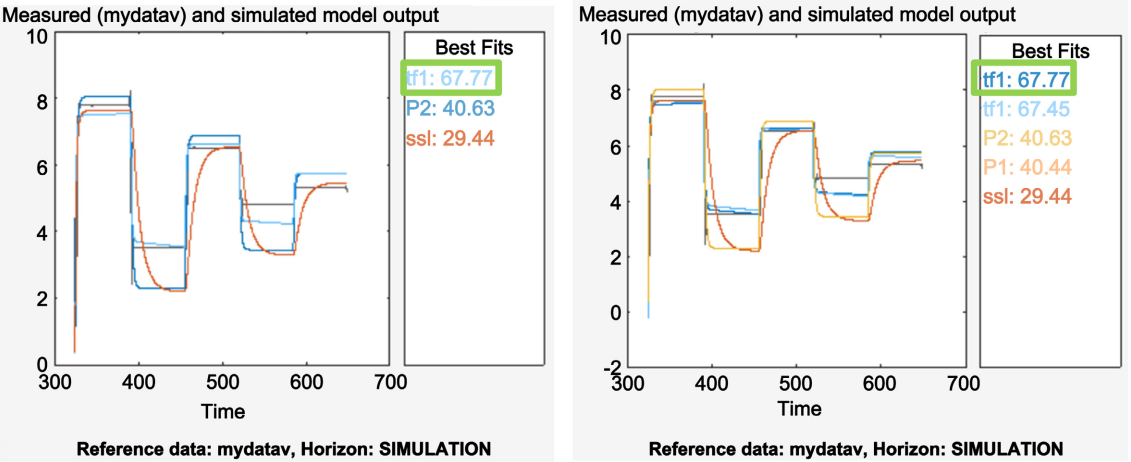
Figure 10 and Figure 11 present the Best-fit values for each tank across all three models, highlighting the strengths and limitations of each approach in representing the system dynamics.



(a) Model Output: T1

(b) Model Output: T2

Figure 10. Comparison of model output for lower tank: T1 and T2.



(a) Model Output: T3

(b) Model Output: T4

Figure 11. Comparison of model output for upper tank: T3 and T4.

3.2. Comparative Analysis of Modeling Structures

This subsection compares the effectiveness of the TF, SS, and PM in capturing the dynamics of the MQTP system. The models were evaluated based on their Best-fit values, which reflect their accuracy in predicting system behavior using validation data. **Figure 12** represents the comparative analysis for these three model structures.

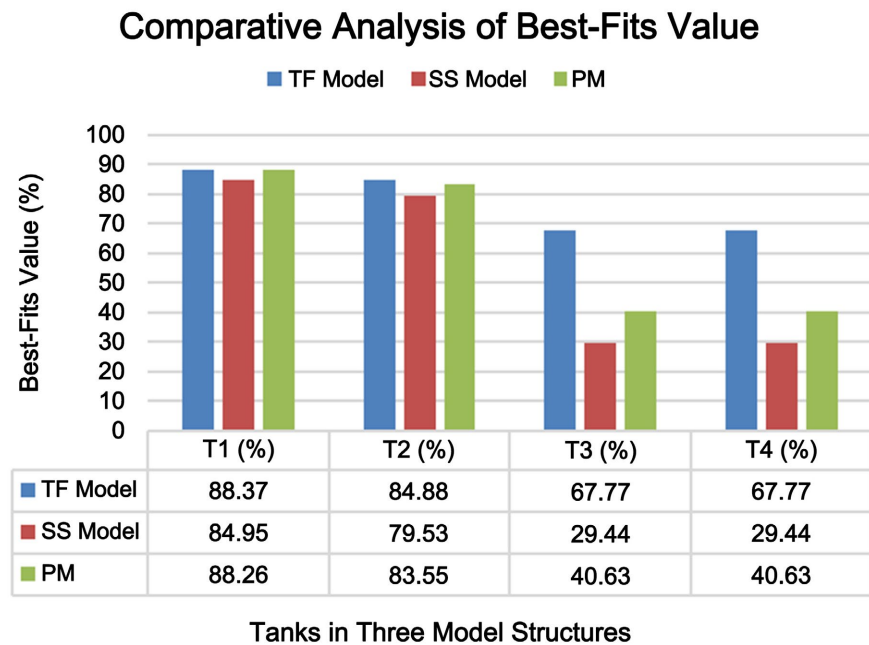


Figure 12. Comparative analysis of TF, SS, and PM in model structures for MQTP.

The TF model provides a straightforward representation of the system's input-output relationship. It is particularly effective for modeling simpler systems or systems with relatively predictable dynamics. For the MQTP system, the TF model demonstrated strong performance for the lower tanks (T1 and T2), where the system dynamics are less complex. However, the TF model's performance was weaker for the upper tanks (T3 and T4) due to the complex nonlinearities and cross-coupling effects between the tanks. Despite this, the TF model remains a valuable choice for applications where computational simplicity is prioritized over capturing complex interactions.

The SS model captures the system's internal dynamics by describing its state variables and interactions through differential equations. It is particularly effective for modeling multivariable systems like the MQTP, where multiple variables (such as tank levels and pump operations) interact. The SS model performed well for the lower tanks, but its performance for the upper tanks was lower than expected. This suggests that while the SS model can model the internal state dynamics, it may require further refinement to accurately capture the full complexity of the upper tank dynamics, where nonlinearities and cross-coupling are more pronounced.

The PM approach is based on simplified physical relationships, such as gains, time constants, and delays. While the PM performed well for the lower tanks, its performance for the upper tanks was significantly lower. These limitations arise from the PM's inability to account for the complex interactions and nonlinear behavior between the tanks. Although the PM model theoretically is computationally efficient, it is unsuitable for systems with significantly nonlinear dynamics or complex interactions between variables.

Among the three identified models, the TF models was selected for closed-loop validation due to its superior balance of accuracy and simplicity. It demonstrated the highest Best-fit values for the lower tanks and acceptable performance for the upper tanks, making it the most practical choice for closed-loop validation [19] [20].

Table 2 shows the Best-fit values and the TF models for the lower and upper tanks. The Best-fit analysis from Section 3.1 reveals that the TF model reliably represents the system's behavior for the lower tanks. Its computation efficiency suits systems that prioritize performance with minimal computational overhead.

Table 2. Best-fit percentage and TF model for each tank in MQTP.

Tanks	Best-fit percentage and TF model	
	Best-fit (%)	Model structures
T1	88.37	$\frac{0.1318s + 0.0404}{s + 0.0399}$
T2	84.88	$\frac{0.2674s + 0.0476}{s + 0.0489}$
T3	67.77	$\frac{0.0159s + 0.0001}{s^2 + 0.0535s + 0.0003}$
T4	67.77	$\frac{0.0612s + 0.0076}{s^2 + 0.308s + 0.0262}$

3.3. Closed-Loop Validation with Experimental Data

The results from the closed-loop validation demonstrated how accurately the identified TF model could predict the behavior of the MQTP system. NRMSE and MAPE values were calculated for each tank, and the simulated output from Simulink was compared to the real plant data. **Figure 13** and **Figure 14** show the findings, which show how well the model replicated the system dynamics and the level of accuracy in predicting the liquid levels for each tank under the same set-point conditions.

Table 3 provides a detailed comparison of each tank's MAPE and NRMSE values, highlighting the TF model's performance within the closed-loop system. The results show that the TF model tracks liquid levels effectively, with T1 and T2 having the lowest error metrics. Specifically, T1 has a MAPE of 10.05%, while T2 has an even lower MAPE of 9.13%. In contrast, T3 and T4 have slightly higher error rates, with MAPE values of 14.97% and 15.00%, respectively. Although these values are higher than those of the lower tanks, they remain within an acceptable

range for control applications. While no universal threshold for MAPE exists, values below 20% are commonly considered acceptable in energy systems and process control applications, as demonstrated in previous studies such as [9] and [21]. Therefore, adopting a $MAPE \leq 16\%$ threshold in this study aligns with established practices in control system modeling. The NRMSE for T1 and T2 are 0.37 and 0.39, respectively, while T3 and T4 have slightly better NRMSE values of 0.34 and 0.35.

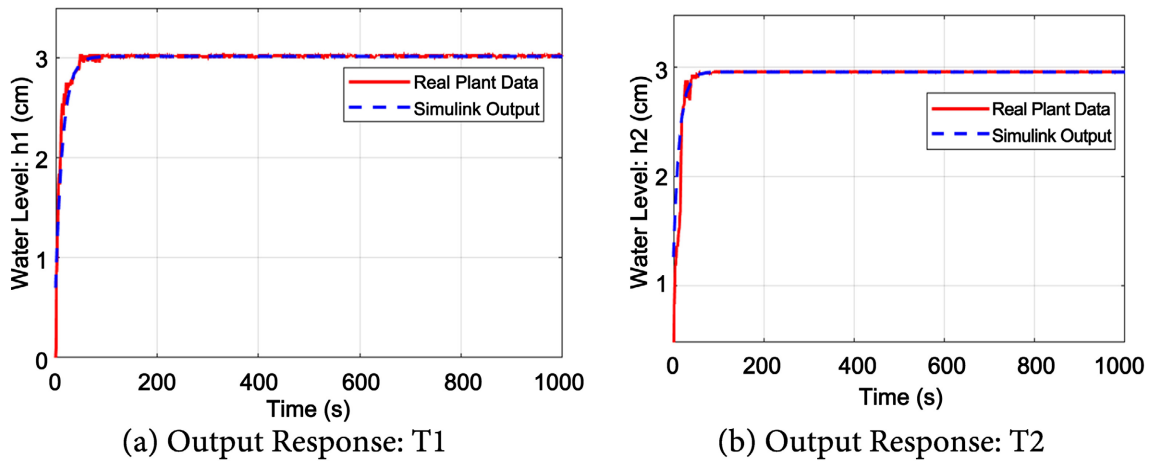


Figure 13. Comparative analysis of simulated and real plant for lower tank.

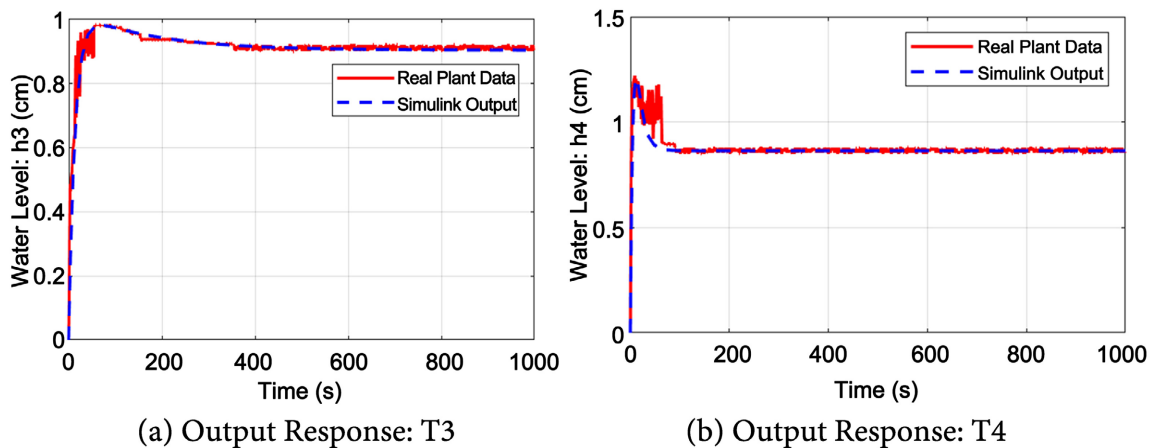


Figure 14. Comparative analysis of simulated and real plant for upper tank.

Table 3. Best-fit percentage and TF model for each tank in MQTPP.

Tanks	Parameter metrics	
	MAPE (%)	NRMSE
T1	10.05	0.37
T2	9.13	0.39
T3	14.97	0.34
T4	15.00	0.35

This closed-loop validation confirms that the TF model is a reliable tool for simulating the MQTP system, particularly for the lower tanks, and provides a solid foundation for real-time applications and further system analysis.

3.4. Implications of Data-Driven Modeling for Energy-Efficient Control

Accurate system modeling is a foundational requirement for designing energy efficient control strategies, particularly in applications where fluid dynamics directly influence performance, stability and energy utilization. The findings of this study underscore the critical role of data driven modeling in enabling precise control of fluid-based energy systems such as thermal energy storage units, microgrids, distributed energy resources (DERs) and water treatment facilities.

Among the evaluated techniques, the Transfer Function (TF) model offers a computationally efficient solution for systems with relatively simple dynamics. Its strong performance in modeling lower tank behaviors suggests suitability for applications like thermal energy storage where fluid flow management is essential for optimizing heat retention and release. The TF model's real time prediction capabilities and low computational overhead make it particularly advantageous for control systems requiring fast response and minimal latency.

In energy storage systems such as pumped hydro or thermal reservoirs, accurate prediction of fluid levels enables more effective scheduling of charge and discharge cycles, thereby improving energy utilization and system stability. The State Space (SS) model, with further refinement, holds promise for more complex energy environments like microgrids and DERs where dynamic interactions between generation, storage and demand must be managed with precision.

Beyond control, accurate data driven models support predictive maintenance by forecasting anomalies in fluid flow or storage levels. This capability enhances system resilience, reduces unplanned downtime and contributes to long term energy efficiency. The TF model's simplicity and responsiveness also make it well suited for real time energy management in buildings and small-scale networks where rapid decision making can significantly reduce energy waste.

In summary, the modeling accuracy achieved through data driven system identification directly impacts the effectiveness of energy control strategies. By enabling precise, responsive and computationally efficient control, these models contribute to the broader goal of sustainable and resilient energy infrastructure.

4. Conclusion

This study has demonstrated that accurate data-driven modeling and system identification are essential precursors to designing energy-efficient control strategies for fluid-based systems such as the Modified Quadruple Tank Process (MQTP). By evaluating Transfer Function (TF), State Space (SS) and Polynomial Model (PM) techniques, the research highlights how model fidelity directly influences control precision, responsiveness and energy optimization. The TF model emerged as the most effective approach, particularly for the lower tanks, achieving high Best-Fit

values while maintaining computational efficiency. Its simplicity and real-time prediction capabilities make it highly suitable for control systems requiring fast actuation and minimal processing overhead, typical in thermal energy storage and water treatment applications. Closed-loop validation using NRMSE and MAPE metrics confirmed that the TF model can be integrated into real-time control architectures, enabling precise tracking of fluid levels and rapid response to setpoint changes. To demonstrate this capability, the validation was conducted using a single setpoint change under controlled conditions. This approach was chosen to verify that, despite lower Best-fit values for the upper tanks, the TF model still performs adequately when evaluated using MAPE and NRMSE metrics. The results support the model's reliability for control applications, particularly in the lower tanks, and show that even the upper tanks remain within acceptable error thresholds. This level of accuracy is critical for minimizing energy waste optimizing charge discharge cycles and maintaining system stability under dynamic operating conditions. The key contribution of this study lies in establishing a clear link between modeling accuracy and energy-efficient control performance. By quantifying the tradeoffs between model complexity, computational load, and predictive reliability, the research provides a practical framework for selecting modeling techniques that align with control objectives in energy systems. This knowledge supports the development of intelligent control designs that operate with high efficiency, resilience, and adaptability. A broader validation involving multiple setpoints and disturbances in MIMO scenarios is planned for future work to further assess model robustness and generalizability. Additionally, hybrid modeling approaches that combine the strengths of TF and SS models could be explored to address systems with mixed dynamic characteristics. Additionally, integrating adaptive control and fault-tolerant features would further enhance the applicability of these models in smart energy networks and sustainable infrastructure.

Acknowledgements

This research was funded by a grant from Ministry of Higher Education of Malaysia (MOHE) through the Fundamental Research Grant Scheme (FRGS), No. FRGS/1/2022/TK07/UTEM/02/09. We would also like to express our sincere gratitude to Universiti Teknikal Malaysia Melaka (UTeM), the Faculty of Electrical Technology and Engineering (FTKE), the Centre of Robotic and Industrial Automation (CERIA) and the Centre of Research and Innovation Management (CRIM) for supporting and contributing to this research.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Zhang, D., Del Rio-Chanona, E.A., Petsagkourakis, P. and Wagner, J. (2019) Hybrid Physics-Based and Data-Driven Modeling for Bioprocess Online Simulation and Optimization. *Biotechnology and Bioengineering*, **116**, 2919-2930.

- <https://doi.org/10.1002/bit.27120>
- [2] Habib, M.K., Ayankoso, S.A. and Nagata, F. (2021) Data-Driven Modeling: Concept, Techniques, Challenges and a Case Study. 2021 *IEEE International Conference on Mechatronics and Automation (ICMA)*, Takamatsu, 8-11 August 2021, 1000-1007. <https://doi.org/10.1109/icma52036.2021.9512658>
 - [3] Bousdekis, A., Lepenioti, K., Apostolou, D. and Mentzas, G. (2021) A Review of Data-Driven Decision-Making Methods for Industry 4.0 Maintenance Applications. *Electronics*, **10**, Article No. 828. <https://doi.org/10.3390/electronics10070828>
 - [4] Patel, H.R. and Shah, V.A. (2019) Decentralized Stable and Robust Fault-Tolerant PI Plus Fuzzy Control of MIMO Systems: A Quadruple Tank Case Study. *International Journal on Smart Sensing and Intelligent Systems*, **12**, 1-20. <https://doi.org/10.21307/ijssis-2019-004>
 - [5] Aranda-Cetraro, I., Pérez-Zúñiga, G., Rivas-Pérez, R. and Sotomayor-Moriano, J. (2023) Nonlinear Robust Control by a Modulating-Function-Based Backstepping Super-Twisting Controller for a Quadruple Tank System. *Sensors*, **23**, Article No. 5222. <https://doi.org/10.3390/s23115222>
 - [6] Kasiyanto, I., Firdaus, H., Lailiyah, Q., Kusnandar, N. and Supono, I. (2024) Advanced Control for Quadruple Tank Process. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, **10**, 1-16. <https://doi.org/10.26555/jiteki.v10i1.27124>
 - [7] Sun, L. and You, F. (2021) Machine Learning and Data-Driven Techniques for the Control of Smart Power Generation Systems: An Uncertainty Handling Perspective. *Engineering*, **7**, 1239-1247. <https://doi.org/10.1016/j.eng.2021.04.020>
 - [8] Ahangar, P.A., Lone, S.A. and Gupta, N. (2023) Combining Data-Driven and Model-Driven Approaches for Optimal Distributed Control of Standalone Microgrid. *Sustainability*, **15**, Article No. 12286. <https://doi.org/10.3390/su151612286>
 - [9] Wang, Q., Li, F., Tang, Y. and Xu, Y. (2019) Integrating Model-Driven and Data-Driven Methods for Power System Frequency Stability Assessment and Control. *IEEE Transactions on Power Systems*, **34**, 4557-4568. <https://doi.org/10.1109/tpwrs.2019.2919522>
 - [10] Pinto, G., Deltetto, D. and Capozzoli, A. (2021) Data-Driven District Energy Management with Surrogate Models and Deep Reinforcement Learning. *Applied Energy*, **304**, Article ID: 117642. <https://doi.org/10.1016/j.apenergy.2021.117642>
 - [11] Araújo, Í.B.Q., Cavalcante, G.B., Lúcio, Y.L.S. and Araújo, F.M.U. (2019) Non-Linear MIMO System Identification System Using the NARX Model. *Electrical Engineering Review*, **95**, 66-72.
 - [12] Chen, Y., Song, L., Liu, Y., Yang, L. and Li, D. (2020) A Review of the Artificial Neural Network Models for Water Quality Prediction. *Applied Sciences*, **10**, Article No. 5776. <https://doi.org/10.3390/app10175776>
 - [13] Wang, X., Pan, M. and Xia, C. (2019) Water Level Data Preprocessing Method Based on Savitzky-Golay Filter. In: *Proceedings of the 2019 International Conference on Modeling, Simulation and Big Data Analysis (MSBDA 2019)*, Atlantis Press, 288-294. <https://doi.org/10.2991/msbda-19.2019.44>
 - [14] Özkan, S., Kara, T. and Arıcı, M. (2017) Modelling, Simulation and Control of Quadruple Tank Process. *Proceedings of the 2017 International Conference on Control, Automation and Systems (ICCAS)*, Bursa, 30 November-2 December 2017, 866-870.
 - [15] A, A., Sriram, S.R. and A, J.L. (2025) Quadruple Spherical Tank Systems with Automatic Level Control Applications Using Fuzzy Deep Neural Sliding Mode FOPID Controller. *Journal of Engineering Research*, **13**, 68-83.

- <https://doi.org/10.1016/j.jer.2023.09.022>
- [16] Naidoo, C. and Pillay, N. (2023) A Data-Driven Approach to Process Control Modelling Using Machine Learning. *Proceedings of the 2023 International Conference on Electronics, Communication and Embedded Systems (ICECET)*, Cape Town, 16-17 November 2023, 1-5.
 - [17] Maier, H.R., Zheng, F., Gupta, H., Chen, J., Mai, J., Savic, D., *et al.* (2023) On How Data Are Partitioned in Model Development and Evaluation: Confronting the Elephant in the Room to Enhance Model Generalization. *Environmental Modelling & Software*, **167**, Article ID: 105779. <https://doi.org/10.1016/j.envsoft.2023.105779>
 - [18] Rajasekhar, N., Nagappan, K.K., Radhakrishnan, T.K. and Samsudeen, N. (2023) Application of Recurrent Neural Networks for Modeling and Control of a Quadruple-tank System. *Advanced Control for Applications*, **6**, e158. <https://doi.org/10.1002/adc2.158>
 - [19] Ljung, L. (1999) *System Identification: Theory for the User*. 2nd Edition, Prentice Hall.
 - [20] Lobontiu, N. (2018) *System Dynamics for Engineering Students: Concepts and Applications*. 2nd Edition, Elsevier.
 - [21] de Myttenaere, A., Golden, B., Le Grand, B. and Rossi, F. (2016) Mean Absolute Percentage Error for Regression Models. *Neurocomputing*, **192**, 38-48. <https://doi.org/10.1016/j.neucom.2015.12.114>