Reinforcement Learning-Based Supervisory PID Tuning for a Quadruple-Tank Process: Experimental Assessment

Mariana Reis Oliveira [a], Amanda Caetano Alvarenga [b], Sergio Mauro da Silva Neiro [a], Luis Cláudio Oliveira-Lopes [a], Fernando V. Lima [c], Thiago Vaz da Costa [a*]

[a] *School of Chemical Engineering, Federal University of Uberlândia (UFU), Uberlândia, Brazil*
[b] *School of Electrical Engineering, Federal University of Uberlândia (UFU), Uberlândia, Brazil*
[c] *Department of Chemical and Biomedical Engineering, West Virginia University, Morgantown, USA*
**thiagocosta@ufu.br*

ABSTRACT

This work investigates the use of Reinforcement Learning (RL) as a supervisory layer for process control, focusing on the online adaptation of controller parameters in complex and nonlinear systems. In this context, PID controller tuning via Q-learning (RL-PID) is employed as a representative case study. The methodology consisted of training a supervisory agent in simulation and validating the learned policy experimentally in a quadruple-tank process using an Arduino/Python platform. Controller performance was assessed using the IAE, ISE, and ITAE indices, comparing the RL-PID strategy against classical IMC and ITAE tuning methods. The results showed that the RL-based supervisory architecture outperformed conventional fixed tuning approaches in terms of cumulative error reduction and mitigation of loop interaction effects. In contrast to the classical ITAE tuning, the proposed method avoided actuator saturation, thereby preserving the mechanical integrity of the pumps. These findings indicate that RL-based supervisory control is a practical and promising approach for intelligent automation, with strong potential for application in multivariable industrial processes that require continuous adaptation under complex operating conditions. Furthermore, the proposed framework can be extended to applications such as Model Predictive Control (MPC) tuning, operating mode selection, constraint management, and contingency handling.

Keywords: Process control; PID control; Reinforcement Learning; Q-Learning; Quadruple-tank process.

1. Introduction

Reinforcement Learning (RL) has emerged as a promising framework for supervisory control in complex chemical processes due to its ability to learn optimal decision policies directly through interaction with the environment. Unlike conventional optimization methods that rely on explicit analytical formulations and fixed assumptions, RL can continuously adapt to process nonlinearities, disturbances, model mismatches, and changing operating conditions. This interaction-based learning capability allows RL agents to perform tasks that require knowledge and experience acquired directly from the process (Shin et al., 2019).

In chemical industries, maintaining process variables at desired operating conditions is critical for ensuring product quality and operational safety. PID control remains the most widely used control strategy in this context; however, conventional tuning methods typically result in fixed controller parameters, which may lead to instability or poor performance in highly nonlinear systems (Seborg et al., 2011). This limitation motivates the development of data-driven approaches for improving controller tuning and closed-loop performance. In this context, RL is particularly attractive when employed as a high-level decision-making layer operating above conventional regulatory controllers, such as Proportional–Integral–Derivative (PID) controllers.

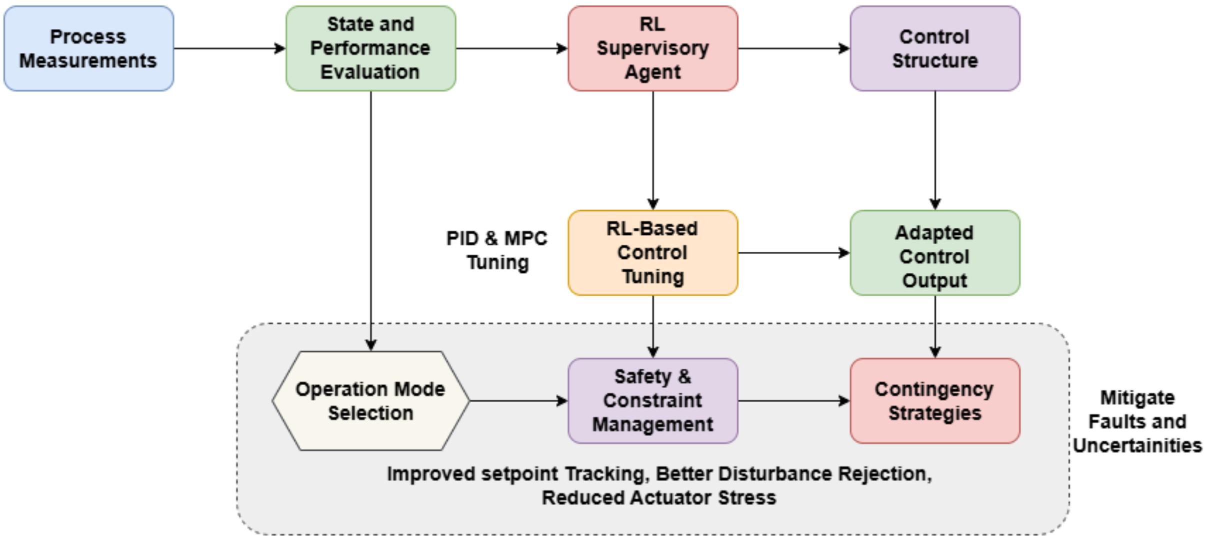
Recent developments in Reinforcement Learning (RL) have broadened its application in process control, particularly for adaptive controller tuning in nonlinear and multivariable systems. Prior research has shown that RL and deep RL strategies enhance tracking performance and robustness, while addressing limitations inherent in classical tuning methods and traditional Q-learning implementations (Yeh and Yang, 2021; Mate et al., 2023; Bujgoi and Sendrescu, 2025). Beyond PID tuning, RL is increasingly explored as a supervisory layer that incorporates safety constraints, actuator limitations, and adaptive optimization within advanced process control architectures. Nevertheless, although simulation-based studies are abundant, experimental validation in real chemical process systems remains limited.

Accordingly, this work focuses on RL-based PID tuning as a representative case study within a broader supervisory framework. The proposed method is implemented and validated using the classical quadruple-tank benchmark process introduced by Johansson (2000) and a low-cost real-time control platform adapted from Pedroso and Batista (2022). In the proposed architecture, the RL agent does not directly manipulate the process. Instead, it is employed for offline RL-based controller tuning in the presence of nonlinear and multivariable process interactions. The performance of the RL-based PI strategy is compared with conventional Internal Model Control (IMC) and Integral of Time-weighted Absolute Error (ITAE) tuning methods to evaluate improvements in tracking performance, mitigation of loop interactions, and reduction of excessive actuator effort under realistic operating conditions.

2. Methodology

Figure 1 illustrates a general supervisory architecture in which Reinforcement Learning (RL) operates as a high-level decision-making layer for complex systems and industrial process control applications. The proposed RL-based supervisory framework provides a general methodology for intelligent process control, with PID tuning representing only one possible application. As illustrated, the architecture considers the acquisition of process measurements, which are used to evaluate the current operating conditions and control performance of the plant. Based on this information, the RL supervisory agent determines appropriate tuning and supervisory decisions according to the process state. By integrating data-driven control adjustment, safety management, and contingency handling into a unified framework, the illustrated architecture can be extended to a wide range of industrial systems requiring continuous performance improvement under complex and uncertain operating conditions.

Figure 1 – RL-based supervisory architecture for complex systems and industrial control applications



Within this supervisory architecture, RL does not directly manipulate the process actuators. Instead, it serves as a supervisory mechanism responsible for determining controller tuning parameters according to the plant operating conditions. In the present study, PID tuning is adopted as a representative case study, in which the RL agent determines the controller parameters (K_p, τ_i, τ_d) to improve setpoint tracking and disturbance rejection. The same framework can be extended to other control structures, such as MPC, through the adjustment of prediction and control horizons, weighting factors, and additional tuning parameters.

The experimental setup used in this study is depicted in Figure 2. It comprises an interconnected quadruple-tank system based on the physical configuration proposed by Pedroso and Batista (2022), in which the control objective is to regulate the liquid levels of the lower tanks. The apparatus includes four cylindrical tanks and two pumps actuated by RS-385 DC motors and controlled via PWM signals from an Arduino Uno microcontroller with 10-bit ADC resolution. The instrumentation used in this work includes MPX5010DP differential pressure sensors for level measurement and YF-S401 flow sensors for flow rate monitoring. Communication between the plant and the Python-based supervisory environment was established via the Modbus TCP/IP protocol with a sampling period of 1.0 s.

Figure 2 – Experimental setup used in this study.

The model parameters were initially determined from preliminary characterization tests and subsequently refined through nonlinear parameter optimization based on the minimization of prediction errors between the nonlinear model and experimental data. The resulting nonlinear model was then employed for RL training in the simulation environment, ensuring a high degree of agreement with the dynamics of the physical plant. Simplified models obtained from experimental step-response tests were used for PI controller tuning based on standard tuning correlations, serving as a basis for comparison. The collected level data were approximated by First-Order Plus Dead Time (FOPDT) models resulting in the transfer function matrix $\boldsymbol{G}(s)$ presented in Equation (1).

$$\boldsymbol{G}(s) = \begin{bmatrix} \dfrac{0.25}{72.77s + 1} e^{-4.30s} & \dfrac{0.16}{81.36s + 1} e^{-7.00s} \\ \dfrac{0.08}{61.18s + 1} e^{-4.80s} & \dfrac{0.32}{77.56s + 1} e^{-3.90s} \end{bmatrix} \quad (1)$$

The Relative Gain Array (RGA) matrix was computed to assess the input–output pairing of the system. The results indicate a diagonal pairing, with Pump 1 paired with Level 1 and Pump 2 with Level 2. Using this configuration, the reference tuning parameters for the classical controllers were determined using the Internal Model Control (IMC) method, considering two values of τ_c, the IMC filter time constant, as well as the minimum ITAE tuning criterion for servo problems. The resulting tuning parameters are presented in Table 1.

Table 1 – PI tuning parameters obtained

Method	Loop	K_c (cm/%)	τ_i (s)
IMC ($\tau_c = \tau/2$)	1	7.21	72.77
	2	5.69	77.56
IMC ($\tau_c = \tau/4$)	1	13.04	72.77
	2	10.43	77.56
ITAE (servo)	1	51.89	71.32
	2	46.80	75.91

2.1 *RL-PID* Architecture.

In this study, the RL-based PID structure is implemented in a supervisory configuration to determine the Kc and τi parameters of a PI controller based on a policy (i.e., a mapping from system states to controller parameters) obtained through prior training. The agent acts every 10 s, which is slower than the control loop's 1.0-second sampling time. The RL formulation is defined by a discretized state space based on the tracking error, an action space associated with incremental adjustments of the PID parameters, and a reward function R_t. The state space is constructed from discretized ranges of the control error $|e|$, divided into predefined intervals, resulting in 18 states per agent. Two independent agents were implemented, each responsible for one control loop, with the interaction between loops treated as a disturbance.

The agent can intervene in the system via a discrete set of five actions that apply fixed increments to the current PI parameters. Upper and lower bounds were imposed on K_c and τ_i to prevent instability during the exploration phase. The reward function R_t (Equation 2), evaluated at each time step t, was designed to promote accurate setpoint tracking while discouraging excessive control effort and unsafe operating conditions. Specifically, the reward function penalizes both the absolute tracking error ($|e|$) and the control effort variation (Δu).

$$R_t = \begin{cases} -500, & if\ h_i \geq 28\ cm\ or\ u_i \geq 100\% \\ +10, & if\ |e_t| < 0.1\ and\ |e_{t-1}| \geq 0.1 \\ +5, & if\ |e_t| < 0.1\ and\ |e_{t-1}| < 0.1 \\ -(|e_t| + 0.1|\Delta u_i|), & otherwise \end{cases} \quad (2)$$

To accelerate convergence, a reward bonus of +10 was assigned for errors close to the setpoint ($|e| < 0.1$ cm), while severe penalties of -500 are imposed to prevent overflow conditions ($h > 28$ cm) and sustained actuator saturation. A weighting factor 0.1 was introduced to ensure that tracking performance remains the primary objective, while still penalizing excessive actuator movement.

3. Results

The RL agent training in simulation lasted approximately 2.5 hours over 10,000 episodes. Figure 3 presents the cumulative reward and cumulative error obtained for each agent. The convergence of the learned policy can be observed through the stabilization of the cumulative reward at values close to 27,000 after approximately 1,500 episodes, while the average error stabilized around 0.02 cm. For implementation in the physical plant, fixed PI parameters were determined as the average values obtained over the last 50 seconds of the training simulation ($K_{c1} = 16.35$, $\tau_{i1} = 49.86$, $K_{c2} = 11.26$ and $\tau_{i2} = 41.15$).

Figure 3 – Cumulative reward and Cumulative error during agent training

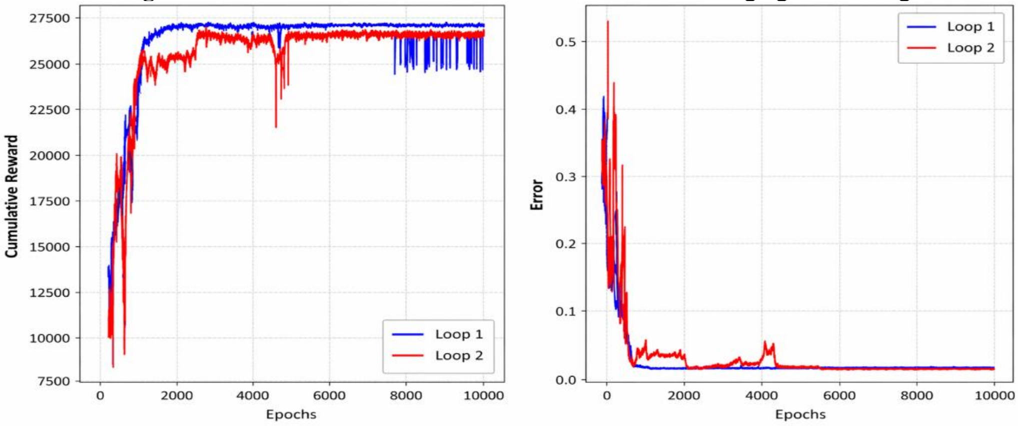


Figure 4 illustrates the comparison between the experimental and simulated system responses using the controller parameters obtained during the training stage. The observed behavior is consistent with the stability achieved in simulation, supporting the effectiveness of the proposed RL-based tuning strategy. Minor discrepancies in the control effort (PWM) relative to the simulated responses can be attributed to uncertainties in the phenomenological model, which does not account for nonlinear head losses or variations in actuator efficiency, as well as to the use of fixed parameters obtained from the training phase for implementation in the microcontroller.

The experimental evaluation of the control strategies was carried out through setpoint changes in the physical quadruple-tank plant. Figure 5 presents a comparison of the closed-loop responses of the lower tank levels for the classical tuning methods and the RL-based tuning strategy. The IMC tuning strategies, although stable, exhibited slow and conservative responses. In contrast, the minimum classical ITAE tuning produced the shortest rise times, but at the expense of large overshoots and oscillatory behavior before reaching steady-state conditions. The RL-based tuning strategy achieved an improved trade-off between response speed and control effort, exhibiting faster response dynamics than the IMC method while requiring significantly smoother control actions than the ITAE-based tuning.

Figure 4 – Simulated vs. Experimental performance behavior

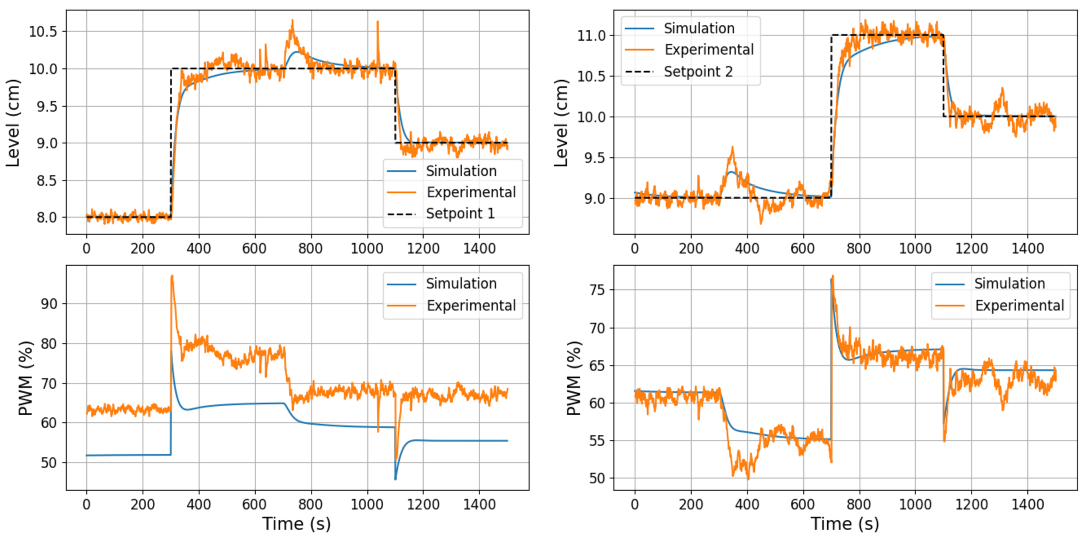
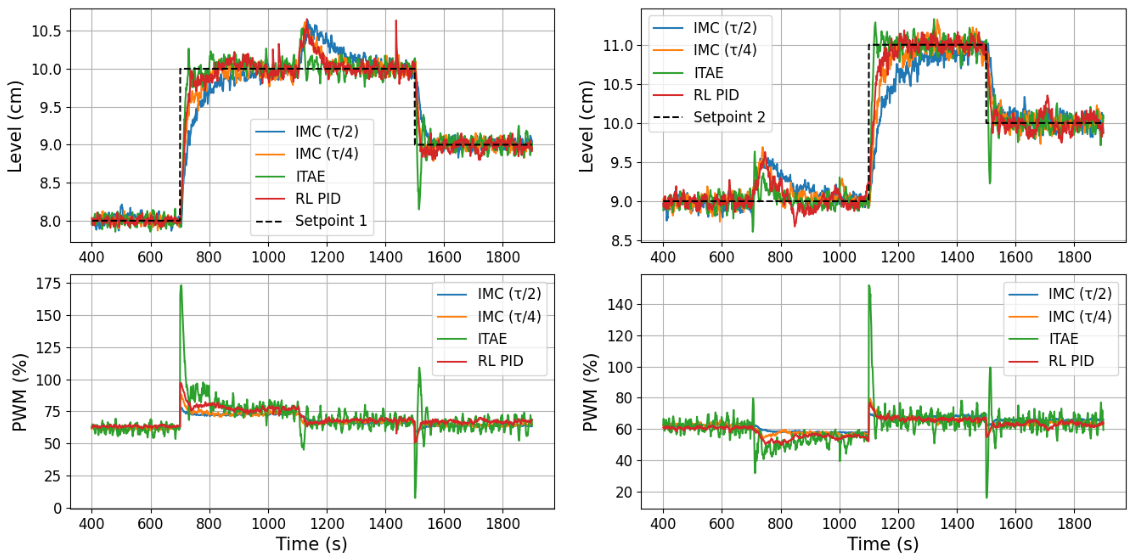


Figure 5 – Closed-loop performance comparison between classical tuning methods and RL-PID.



To quantitatively compare the different techniques, the IAE, ISE, and ITAE performance indices were calculated for each method and are presented in Table 2. The analysis confirms the superior performance of the RL-based tuning strategy in the simulation environment, where it achieved the lowest IAE, ISE, and ITAE values among all evaluated methods. This result is expected, as the simulation environment does not account for measurement noise and external disturbances present in the experimental setup.

Under real operating conditions, the effectiveness of the proposed strategy is further evidenced by its performance relative to classical approaches: the RL-based controller outperformed the IMC tuning methods across all performance indices. Although the ITAE-based tuning achieved lower error indices in some cases, this performance was obtained at the expense of excessive control effort, leading to actuator saturation and increased sensitivity to measurement noise due to the higher proportional gains.

Table 2 – Performance indices normalized using RL-PI values for all controllers analyzed

	Loop	PI IMC ($\tau_c = \tau/2$)	PI IMC ($\tau_c = \tau/4$)	PI ITAE (servo)	RL-PI (simulation)	RL-PI (experimental)
IAE	1	1.99	1.85	1.72	1.00	1.84
	2	1.93	1.84	1.56	1.00	1.57
ISE	1	3.06	3.07	2.99	1.00	3.06
	2	3.69	3.85	3.34	1.00	3.27
ITAE	1	2.21	1.71	1.59	1.00	1.69
	2	2.02	1.53	1.15	1.00	1.29

In contrast, the RL-based strategy achieved a more balanced performance, maintaining satisfactory tracking while avoiding excessive control action. This improvement is mainly attributed to the flexibility of the reward function, which enables a multidimensional search for controller parameters without being restricted to a single tuning variable, such as τ_c in the classical IMC approach.

4. Conclusion

This work demonstrated that the adaptive RL-PID architecture can overcome several limitations of conventional fixed-tuning methods by performing a multidimensional search over the controller parameter space, thereby achieving a more favorable balance between tracking performance and actuator integrity. Unlike the IMC and ITAE approaches, the supervisory agent was able to adapt to the coupled dynamics of the quadruple-tank system while maintaining robust performance in the presence of experimental uncertainties and model-plant mismatch. Although promising, the proposed strategy remains dependent on the quality of the model used during training and does not provide formal stability guarantees during learning, highlighting the need for additional safety mechanisms in critical applications. As future work, investigating Deep Reinforcement Learning algorithms, such as Deep Q-Networks (DQN), is recommended to mitigate the curse of dimensionality in continuous state spaces. In addition, direct comparisons with advanced control techniques, such as Model Predictive Control (MPC), should be explored. Other relevant extensions include the development of SCADA interfaces for real-time supervision and the implementation of direct RL-based control of the PWM signal, enabling a fully autonomous, intelligent industrial control framework.

Acknowledgment: The authors acknowledge the financial support provided by FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) through the PIBIC program (EDITAL DIRPE-UFU No. 1/2024) and the Universal Project (Process No. APQ-02534-24), which enabled this research.

References

D. Seborg, T. Edgar, D. Mellichamp e F. Doyle III: Process Dynamics and Control (3a. Ed.), Wiley: 2011.

G. Bujgoi e D. Sendrescu: Tuning of PID controllers using reinforcement learning for nonlinear system control. Processes (13), 735, 2025.

J. Shin, T. A. Badgwell, K. H. Liu & Lee, J. H. Reinforcement learning–overview of recent progress and implications for process control. Computers & Chemical Engineering (127), 282-294, 2019.

K. H. Johansson: The Quadruple-Tank Process: A Multivariable Laboratory Process with an Adjustable Zero, IEEE Transactions on Control Systems Technology (8), 456–465, 2000.

L. Pedroso e P. Batista: Reproducible Low-Cost Flexible Quadruple-Tank Process Experimental Setup for Control Educators, Practitioners, and Researchers, Journal of Process Control (118), 82–94, 2022.

Y. Yeh e K. Yang: Design and comparison of reinforcement-learning-based time-varying PID controllers with gain-scheduled actions. Machines (9), 319, 2021.

S. Mate, P. Pal, A. Jaiswal e S. Bhartiya: Simultaneous tuning of multiple PID controllers for multivariable systems using deep reinforcement learning. Digital Chemical Engineering (9), 100131, 2023.