A Deep Reinforcement Learning Approach to Slug Flow Control in Oil and Gas Applications

José R. T. Neto a,\*, Bruno D. O. Capron a, Argimiro R. Secchi a,b, Antonio Del Rio Chanonac

aDepartment of Chemical Engineering/School of Chemistry, Federal University of Rio de Janeiro, 149 Athos da Silveira Ramos Avenue, Rio de Janeiro, RJ, 21941-909, Brazil

bChemical Engineering Program/COPPE, Federal University of Rio de Janeiro, 149 Athos da Silveira Ramos Avenue, Rio de Janeiro, RJ, 21941-972, Brazil

cDepartment of Chemical Engineering, Imperial College London, South Kensington Campus, London, SW7 2AZ, United Kingdom

[joseneto@eq.ufrj.br](mailto:joseneto@eq.ufrj.br)

Abstract

In deep water oil extraction, supervisory control is often challenged by the oscillatory nature of slug flow. This work introduces an adaptive anti-slug control mechanism using the Proximal Policy Optimization (PPO) algorithm, a Deep Reinforcement Learning (DRL) approach that dynamically adjusts to the complexities of slug flow. The proposed strategy involves the control of the Permanent Downhole Gauge (PDG) pressure through two variables: choke valve opening and gas lift flow rate, representing real-world extraction challenges. The methodology was based on a simulation-based learning approach, utilizing the Fast Oil and Water Model (FOWM). The work compares a baseline PPO model (PPO-base) with a modified version that incorporates penalties on control actions and for violations of state constraints (PPO-pen). Results show that PPO-pen surpasses PPO-base in key training metrics, indicating that moderate penalties can enhance control effectiveness and give more adherence to operational constraints. Testing also reveals that policies from PPO-pen are more efficient than PPO-base, indicating a strategy that is resource-efficient and effective in maintaining critical operational parameters. Furthermore, PPO-pen achieves a lower integral time-weighted absolute error (ITAE), highlighting its accuracy in control and error minimization. The findings demonstrate the potential of DRL, particularly PPO with calibrated penalty parameters, in refining control strategies for complex processes like deep water oil extraction.

**Keywords:** Multiphase Flow Control, Subsea Control Systems, Operational Stability.

* 1. Introduction

Deep water extraction is increasingly vital for the oil and gas industry as hydrocarbon sources deplete. This process involves extracting hydrocarbons from subsea wells and transporting them to surface platforms via flowlines, manifolds, and risers, followed by export to shore for refining. A primary challenge within the production line is managing slug flow, which can cause significant operational issues, including the overflow of inlet separators in offshore oil fields (Taitel, 1986). For control design, complex process models, like those in the OLGA® simulator (Bendlksen et al., 1991), are computationally expensive. Instead, this work adopts a simplified, yet effective model called Fast Offshore Wells Model (FOWM) (Diehl et al., 2017), effectively capturing the critical dynamics of casing heading and slugging in both terrain and riser. Traditionally, the industry has relied on Proportional-Integral-Derivative (PID) control strategies. Several studies have focused on the effectiveness of PID control in slug flow management, exemplified by works like Di Meglio et al. (2012a), Storkaas and Skogestad (2008), and Jahanshahi (2013). The PID approach, due to its practicality, has been a focal point in research, with significant efforts dedicated to optimizing its tuning for anti-slug control (Pedersen et al., 2014), as seen in the methods proposed by Godhavn et al. (2005). This work introduces a novel approach to slug flow control in oil and gas operations, transitioning from traditional PID controller tuning to the application of Deep Reinforcement Learning (DRL). The primary objective is to explore and establish the effectiveness of DRL as a supervisory control method that can adaptively and dynamically handle the complex, fluctuating conditions characteristic of slug flow. This shift represents a potentially more robust solution for scenarios where traditional control strategies might struggle to maintain stability and efficiency.

* 1. Methodology
     1. Slug Flow Model

FOWM is a composite model that divides the production system into three segments: the reservoir and wellbore, the gas lift, and the production line, which includes both the flowline and riser (Figure 1).

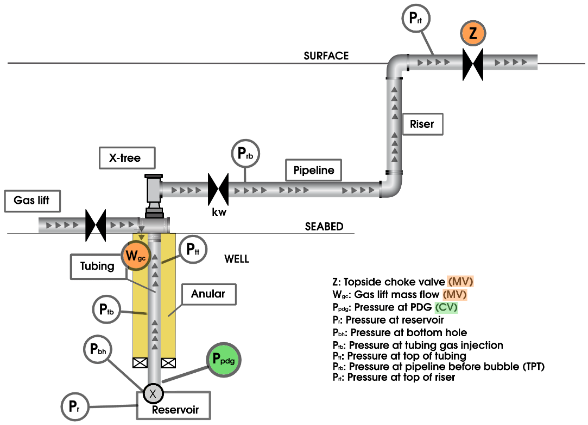


Figure 1: FOWM model for oil and gas production simulation (Adapted from Diehl et al., 2017).

This trisection is articulated through a system of ordinary differential equations (ODEs), based on conservation of mass equations that represent the masses of gas and liquid in different sections of the system. The parameter set from Rodrigues et al. (2018) was selected for characterizing FOWM in this work due to its good alignment with the specific region and initial conditions under consideration, particularly in cases of severe slugging.

* + 1. Proximal Policy Optimization

This work employs the Proximal Policy Optimization (PPO) algorithm (Schulman et al., 2017), an on-policy DRL algorithm known for balancing the complexity of policy gradient methods with the stability of trust region methods. PPO restricts policy updates through a clipped surrogate objective function, creating a trust region to ensure stable and moderate changes in the policy. The algorithm uses both a policy network (actor) and a value network (critic), with the latter aiding in stabilizing training by estimating the value function. The advantage function, , critical for directing the policy towards more beneficial actions, is calculated using Generalized Advantage Estimation (GAE) (Schulman et al., 2015). The general hyperparameters for the PPO implementation can be found on Table 1.

Table 1: PPO Hyperparameters

| Parameter | Value | Definition |
| --- | --- | --- |
| dt | 200 | Simulation time step (seconds) |
| max\_timestep | 4320 | Max time steps per episode (10 days) |
| learning\_rate | 5×10−4 | Learning rate for optimization |
| clip\_range | 0.2 | Range for clipping policy updates |
| clip\_range\_vf | 0.2 | Clipping range for value function updates |
| total\_timesteps | 1×107 | Total training time steps |
| gamma | 0.99 | Discount factor for future rewards |
| gae\_lambda | 0.99 | GAE parameter for bias-variance trade-off |
| ent\_coef | 5×10−3 | Entropy term coefficient in loss calculation |
| batch\_size | 240 | Batch size for gradient updates |
| n\_steps | 480 | Steps collected before each update |
| n\_epochs | 10 | Number of passes over each batch of steps |
| seed | 1990 | Seed for random number generation |

The neural network architecture for the PPO model was determined through a heuristic exploratory process, similar to Lee (2020). Initially, a range of neural network structures and activation functions were trialed, leading to the selection of the Mish activation function (Misra 2020), defined as f, where is . Mish was selected for its stability and superior performance metrics, trained using the Adam optimizer. The final architecture, featuring two fully connected layers with 16 neurons each for both policy (actor) and value (critic) networks, was chosen for its simplicity and computational efficiency. The architecture's effectiveness was validated in deterministic policy evaluations, prioritizing average reward and episode mean length as key performance metrics.

* + 1. Case Study

The study's slugging process control environment was developed using the Gym toolkit (Brockman et al., 2016). The permanent downhole gauge pressure (PDG) was selected as the controlled variable for its precision in capturing well dynamics and its direct impact on production rates. A lower PDG is typically associated with higher production, due to the increased pressure differential with the reservoir. The choke valve position and the gas lift flow rate were selected as manipulated variables. The choke valve adjusts the well's output, while the gas lift enhances oil recovery by lightening the fluid column in the well. The observation space consists of a set of variables critical for assessing the well’s states (Table 2).

Table 2: Observation Space

|  |  |
| --- | --- |
| Parameter | Definition |
|  | The pressure at the top region of the riser |
|  | The pressure at the flowline before the bubble |
|  | The controlled pressure at PDG |
|  | The pressure at the top of the tube |
|  | The pressure at the bottom of the tube |
|  | The pressure at the base of the well |
|  | The mass flow rate of the liquid exiting the system |
|  | The mass flow rate of the gas exiting the system |

The case study was derived by data from a real-world oil and gas well, referred to as well A (Diehl et al., 2017), illustrated by Figure 2. This well was chosen for its stable limit cycles and controllable gas lift system, providing a rich dataset for DRL training.

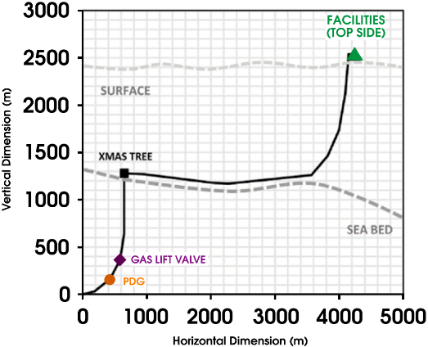


Figure 2: Schematic representation of the case study: well A.

In the control environment, the initial conditions were selected as conditions that could induce slug flow if the system was left uncontrolled. A conservative control strategy, as suggested by Diehl et al. (2018), was adopted, allowing only minor adjustments at each simulation step, of about ±0.048 % for choke valve opening and ±4.762 Sm³/10³ d for gas injection rate. The PDG setpoint was systematically varied over four stages within each episode, aligned with the episode's length. Each stage spanned a quarter of the maximum time steps, with the setpoint starting at 210 bar and gradually decreasing to 203 bar. This approach was specifically designed to test the controller's adaptability to setpoint transitions within the slug flow region, starting at an initial setpoint already within the slug flow conditions. In the PPO model, the reward function aimed to maintain the PDG pressure setpoint and ensure control stability, calculating the error between current and reference pressures with integral time-weighted absolute error (ITAE), to emphasize the duration and magnitude of deviations. The reward bonus scaled according to error proximity to three tolerance zones, with higher rewards for smaller deviations. The reward function also incorporates penalties for variance and mean absolute changes in the control actions, aimed at promoting smoother operations. The cumulative reward at each time step is computed using the following formulation:

|  |  |
| --- | --- |
|  | (7) |

In this formula, is determined by the normalized error’s alignment with the tolerance zones, is the variance for each control action, and is the mean absolute change. Furthermore, constraint handling was implemented during training to guide the learning process. This involves terminating a simulation episode and applying a penalty if state variables exceed the predefined bounds. This mechanism, primarily a training tool, encourages the agent to adhere to operational limits.

* 1. Results and Discussion

This section compares two simulation tests: PPO-base with no penalties (, , and set to 0) and PPO-pen with penalties , , and set to (0.5, 0.5, 2.0). Key metrics evaluated were Episode Length Mean and Episode Reward Mean, indicating system stability and alignment with operational objectives, respectively. Training progress and performance of both scenarios are detailed in Figure 3, showing that PPO-pen outperformed PPO-base in both episode length and reward means. This suggests that penalties enhanced operational constraints adherence and led to smoother control actions, demonstrating the effectiveness of penalty-based training in control strategies under severe initial conditions.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated

Figure 3: Training performance comparison between PPO-base and PPO-pen.

A detailed performance analysis of both configurations during a test episode (Figure 4) revealed several key findings. In choke valve control, PPO-pen achieved a higher final value, indicating better production efficiency. In gas lift rate control, PPO-pen demonstrated more efficient resource utilization with a systematic reduction in gas lift, contrasting with PPO-base’s constant rate. For the PDG pressure control, both scenarios performed similarly, but PPO-pen showed superior consistency and stability, evidenced by a lower ITAE.

A screenshot of a graph

Description automatically generated

Figure 4: Comparative analysis of control actions and performance metrics in a test episode.

* 1. Conclusions

This work implemented and evaluated PPO-based control strategy for deep water oil extraction, comparing a baseline model (PPO-base) against a penalty parameter-integrated version (PPO-pen). The results showed that PPO-pen significantly outperformed PPO-base in training metrics, demonstrating the benefits of incorporating moderate penalties in control strategies for balancing operational objectives and constraints adherence. In testing, PPO-pen presented a more resource-efficient control approach, optimizing choke valve position and reducing gas lift usage. Both models effectively were able to maintain the PDG pressure at setpoint and far from slug conditions. Considering the error minimization, PPO-pen achieved lower ITAE, indicating enhanced control performance. These findings highlight the efficacy of DRL, particularly the PPO framework with tailored penalties, in controlling complex systems like deep water oil extraction. They open avenues for further research in applying these methods to varied operational scenarios and integrating more real-world constraints.

References

Y. Taitel, 1986, "Stability of severe slugging," International Journal of Multiphase Flow, Vol. 12, No. 2, pp. 203-217, DOI: 10.1016/0301-9322(86)90026-1.

K.H. Bendlksen, Dag Malnes, Randl Moe, Sven Nuland, 1991, "The Dynamic Two-Fluid Model OLGA: Theory and Application," SPE Production Engineering, Vol. 6, No. 02, pp. 171-180, DOI: 10.2118/19451-pa.

F.C. Diehl, T.K. Anzai, C.S. Almeida, O.F. Von Meien, S. Simões Neto, V.R. Rosa, M.C.M.M. Campos, F. Reolon, G. Gerevini, C. Ranzan, M. Farenzena, J.O. Trierweiler, 2017, "Fast Offshore Wells Model (FOWM): A practical dynamic model for multiphase oil production systems in deep water and ultra-deep water scenarios," Computers & Chemical Engineering, Vol. 99, pp. 304-313, DOI: 10.1016/j.compchemeng.2017.01.036.

F. Di-Meglio, G.O. Kaasa, N. Petit, V. Alstad, 2012, "Model-based control of slugging: Advances and challenges," Proceedings of the 2012 IFAC Workshop on Automatic Control in Offshore Oil and Gas Production, pp. 109–115.

E. Storkaas, S. Skogestad, 2008, "Controllability analysis of two-phase pipeline-riser systems at riser slugging conditions," Control Engineering Practice, Vol. 15, pp. 567–581.

E. Jahanshahi, 2013, "Control Solutions for Multiphase Flow - Linear and nonlinear approaches to anti-slug control," Ph.D. thesis, Norwegian University of Science and Technology, Department of Chemical Engineering, Trondheim, Norway.

S. Pedersen, P. Durdevic, Z. Yang, 2014, "Learning control for riser-slug elimination and production-rate optimization for an offshore oil and gas production process," Proceedings of The 19th World Congress of the International Federation of Automatic Control.

J.M. Godhavn, M.P. Fard, P.H. Fuchs, 2005, "New slug control strategies, tuning rules and experimental results," Journal of Process Control, Vol. 15, pp. 547-557.

R.F. Rodrigues, J.O. Trierweiler, M. Farenzena, 2018, "New methodology for parameter estimation of offshore slug models with Hopf bifurcation," Computers & Chemical Engineering, Vol. 117, pp. 247–255, DOI: 10.1016/j.compchemeng.2018.06.012.

J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, 2017, "Proximal Policy Optimization Algorithms," arXiv:1707.06347, URL: https://arxiv.org/abs/1707.06347.

J. Schulman, P. Moritz, S. Levine, M. Jordan, P. Abbeel, 2015, "High-Dimensional Continuous Control Using Generalized Advantage Estimation," arXiv:1506.02438, URL: https://arxiv.org/abs/1506.02438.

D. Lee, 2020, "Comparison of Reinforcement Learning Activation Functions to Improve the Performance of the Racing Game Learning Agent," Journal of Information Processing Systems, Vol. 16, No. 5, pp. 1074–1082.

Diganta Misra, 2020, "Mish: A Self Regularized Non-Monotonic Activation Function," arXiv preprint arXiv:1908.08681.

G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, W. Zaremba, 2016, "OpenAI Gym," arXiv:1606.01540, URL: https://arxiv.org/abs/1606.01540.

F.C. Diehl, C.S. Almeida, T.K. Anzai, G.Gerevini, S. Simões Neto, O.F. Von Meien, M.C.M.M. Campos, M. Farenzena, J.O. Trierweiler, 2018, "Oil production increase in unstable gas lift systems through nonlinear model predictive control," Journal of Process Control, Vol. 69, pp. 58-69, DOI: 10.1016/j.jprocont.2018.07.009.