Otimização Econômica de Sistemas Gas Lift via Aprendizado por Reforço com Restrição anti-golfadas

Italo Costa Soares [a,*], Daniel Diniz Santana[a], Raony Maia Fontes[a]

[a] *Programa de Recursos Humanos (PRH) da ANP em Descarbonização e Transformação Digital para a Indústria de Petróleo e Gás (PRH41/UFBA), Escola Politécnica, UFBA, Salvador-BA, Brasil*
[*] *isoares@ufba.br*

RESUMO

Sistemas de *gas lift* (GL) estão diretamente associados ao desempenho econômico da produção de petróleo, uma vez que influenciam tanto a eficiência do escoamento quanto os custos operacionais. Um dos principais desafios desses sistemas é a ocorrência de regimes instáveis de escoamento, conhecidos como golfadas (slugging), que comprometem a estabilidade operacional e reduzem a eficiência da produção. Nesse contexto, este trabalho investiga o uso de aprendizado por reforço (RL) para a otimização operacional de poços com elevação artificial por injeção de gás, visando maximizar a produção de óleo com uso eficiente do gás injetado, considerando explicitamente restrições de operabilidade anti-golfadas. O agente proposto é treinado por meio do algoritmo Twin Delayed Deep Deterministic Policy Gradient (TD3), utilizando uma função de recompensa baseada no lucro operacional, definida a partir da receita associada à produção de óleo e dos custos relacionados à injeção de gás. Para lidar com a ocorrência de golfadas, incorpora-se explicitamente uma fronteira de operabilidade à formulação do problema de RL, por meio de um termo de penalização na função de recompensa, restringindo a operação a regiões estáveis do sistema. Os experimentos são conduzidos em ambiente simulado, utilizando um modelo fenomenológico baseado nos estudos de Eikrem et al., 2004, representando um poço de petróleo sujeito a diferentes condições operacionais, incluindo variações na razão gás-óleo (RGO), utilizadas como perturbações do sistema. Os resultados demonstram que o agente de RL é capaz de aprender políticas operacionais que maximizam o lucro do processo ao mesmo tempo em que evitam regiões associadas ao regime de golfadas. Os resultados obtidos evidenciam o potencial da abordagem proposta para incorporar restrições operacionais críticas diretamente ao processo de aprendizado, destacando o RL como uma alternativa promissora para otimização dinâmica de sistemas *gas lift*.

Palavras-chave: *gas lift*; Aprendizado por Reforço, Otimização Dinâmica.

1 Introdução

A redução gradual da pressão do reservatório ao longo da vida produtiva do poço torna necessária a utilização de métodos de elevação artificial para garantir a continuidade da produção. Entre as principais técnicas empregadas na indústria do petróleo, destaca-se o *gas lift*, método baseado na injeção de gás na coluna de produção com o objetivo de reduzir a densidade da mistura óleo-gás e facilitar o escoamento dos fluidos até a superfície.

Entretanto, o aumento excessivo da vazão de gás injetado intensifica as perdas de carga por atrito ao longo da coluna de produção, elevando as dissipações de energia do escoamento e reduzindo a eficiência do levantamento. Como consequência, a produção de óleo passa a decrescer após determinado ponto de operação. Dessa forma, conforme discutido por Cooksey e Pool, 1995, a eficiência econômica de poços operando por *gas lift* depende diretamente da relação entre o volume de óleo produzido e o consumo de gás injetado.

Porém, essa técnica de elevação artificial apresenta o fenômeno de golfadas, caracterizado pela formação de bolhas de gás que induzem oscilações de pressão e vazão no sistema, podendo comprometer a operação e danificar equipamentos. Para evitar regiões de golfada, estratégias de controle tornam-se necessárias. Ao mesmo tempo, a otimização operacional é fundamental para maximizar a produção e o retorno econômico.

Entretanto, em sistemas sujeitos ao regime de golfadas, o comportamento oscilatório persistente invalida, em determinadas condições operacionais, a hipótese de estado estacionário. Isso limita a aplicação de métodos de otimização estacionária, como o Steady-State Real-Time Optimization (SSRTO). Nesse contexto, abordagens de otimização dinâmica tornam-se mais adequadas, uma vez que consideram explicitamente a dinâmica temporal do sistema e seus regimes transientes. Na literatura, trabalhos como de Sousa Santos et al., 2018 exploram a otimização dinâmica direta do processo por meio de programação dinâmica e controle ótimo. Embora tais abordagens

permitam a obtenção de trajetórias ótimas de operação, o modelo utilizado é simplificado e a solução encontrada é em malha aberta, sem considerar explicitamente fenômenos como golfadas, perturbações operacionais ou restrições industriais mais realistas. Por outro lado, Aranha Ribeiro et al., 2024 propõem diferentes estratégias combinando Economic Model Predictive Control (EMPC) e Real-Time Optimization (RTO), incorporando otimização econômica em malha fechada, tratamento explícito de restrições e maior robustez frente a perturbações e incertezas do processo. Contudo, tais abordagens apresentam elevado custo computacional e dificuldades de escalabilidade, especialmente em sistemas de grande porte e alta complexidade dinâmica. O aprendizado por reforço (RL), por sua vez, vem sendo amplamente investigado na literatura como uma alternativa promissora para problemas de otimização e controle de processos. Trabalhos como os de de Rezende Faria et al., 2024 demonstram o uso do RL como ferramenta complementar para compensar ou aprimorar ações de otimizadores em tempo real aplicados a sistemas *gas lift*, proporcionando maior adaptabilidade frente a incertezas e dinâmicas não modeladas do processo. Neste contexto, este trabalho investiga a aplicação de aprendizado por reforço para geração direta das ações de abertura da válvula *choke* e da vazão de gás injetado em um poço *gas lift*, visando otimizar o desempenho operacional do sistema a partir de uma função de recompensa baseada no lucro do campo. Adicionalmente, considera-se uma fronteira de operabilidade capaz de caracterizar regiões de operação estável e de ocorrência de golfadas, permitindo penalizar condições operacionais indesejáveis durante o processo de aprendizado do agente.

2 Fundamentação Teórica

O aprendizado por reforço consiste em treinar um agente para interagir com um ambiente de modo a maximizar a recompensa acumulada ao longo do tempo. Em cada instante t, o agente observa o estado S_t, executa uma ação A_t e recebe uma recompensa R_t, que conduz à transição para o estado S_{t+1}. Esse processo é modelado como um Processo de Decisão de Markov (MDP), cujo objetivo é aprender uma política π que maximize o retorno esperado:

$$J = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t R_t\right], \tag{1}$$

onde $\gamma \in (0,1)$ é o fator de desconto.

Neste trabalho, utiliza-se o algoritmo TD3 (Twin Delayed Deep Deterministic Policy Gradient), proposto por Fujimoto et al., 2018, que tem como objetivo reduzir instabilidades no treinamento e o viés de superestimação da função valor, em comparação a outros algoritmos da mesma família, como o Deep Deterministic Policy Gradient (DDPG) proposto por Lillicrap et al., 2019.

O TD3 é um algoritmo do tipo Actor-Critic, baseado em duas redes neurais artificiais: o ator e o crítico, com política determinística. O ator gera a ação de acordo com a Equação (2):

$$A_t = \pi_\theta(S_t), \tag{2}$$

enquanto o crítico avalia essa ação ao estimar a recompensa acumulada esperada, conforme a Equação (3):

$$Q(S_t, A_t \mid \phi). \tag{3}$$

O principal avanço do TD3 em relação a outros algoritmos Actor-Critic está no uso de dois críticos independentes, treinados simultaneamente. O alvo temporal é calculado a partir do menor valor estimado entre eles, o que reduz o viés de superestimação.

Além disso, a política-alvo é suavizada por meio da adição de um pequeno ruído à ação durante o cálculo do alvo de recompensa, evitando a exploração de picos artificiais na função de valor. Por fim, a atualização do ator ocorre com menor frequência do que a dos críticos, o que contribui para maior estabilidade no aprendizado.

3 Metodologia

3.1 Sistema de elevação gas lift

O modelo escolhido para representar o sistema de elevação artificial e testar o RL como otimizador foi um sistema dinâmico simplificado de *gas lift*, com um poço e atrito, baseado nos trabalhos de Eikrem, Imsland e Foss (2004). (representado na figura 1).

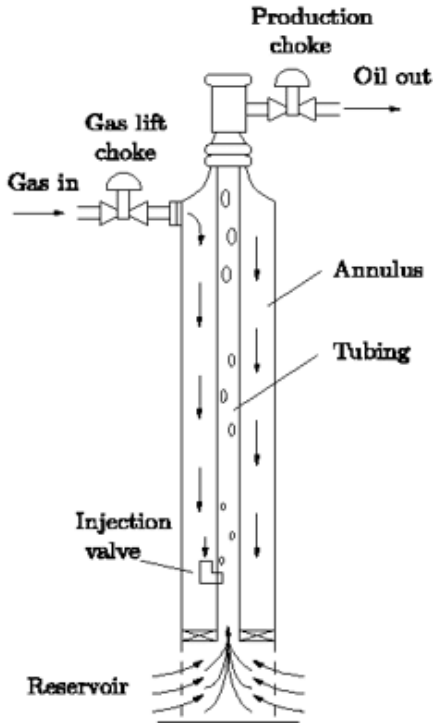

Figura 1: Esquema do sistema de elevação artificial por *gas lift* Eikrem et al., 2004 modificado

Como o sistema é suscetível a golfadas, SOUZA et al., 2025 realizaram uma análise de operabilidade para mapear o comportamento dinâmico do sistema no espaço das variáveis operacionais e identificar regiões associadas aos regimes de operação com e sem ocorrência de instabilidades.

A partir desse mapeamento, foi realizada uma identificação da fronteira de separação entre regimes operacionais, a qual foi aproximada por um modelo linear no espaço de entradas do processo. Essa aproximação resultou na fronteira de operabilidade dada pela equação (4):

$$F = -10.77\,u_1 + 4.61\,w_{\text{inj}} + 76.92\,\text{RGO} + 4.85 \geq 0 \;\Rightarrow\; \text{região estável} \tag{4}$$

Onde u_1 representa a abertura da válvula *choke*, W_{inj} a injeção de gás e *RGO* a razão gás-óleo do processo. A condição $F \geq 0$ define a região operacional associada a comportamento estável, enquanto $F < 0$ delimita a região associada à ocorrência de golfadas. Posteriormente, essa fronteira de estabilidade é utilizada como restrição explícita de operação, sendo incorporada ao *framework* de aprendizado por reforço na forma de um termo de penalidade, com o objetivo de induzir o agente a evitar regiões de golfadas do espaço de controle.

3.2 TD3 como otimizador

O agente de aprendizado por reforço (RL) proposto neste trabalho tem como objetivo determinar a taxa de injeção de gás e a abertura da válvula *choke* do poço, visando maximizar o lucro e, consequentemente, a produção total de óleo, mesmo na presença de distúrbios, como variações na razão gás-óleo (RGO). O treinamento foi realizado a partir de condições iniciais aleatórias dentro da faixa operacional do sistema, com taxas de injeção de gás variando de 0 a 3 kg/s.

Para que o agente atue como um otimizador, o primeiro passo consiste na definição do ambiente em que será treinado, bem como das variáveis observáveis e da forma de interação com esse ambiente. Para a construção do ambiente, utilizou-se a biblioteca *Gymnasium*, adotando-se o sistema apresentado previamente.

As variáveis observadas (s_t) pelo agente são a pressão de fundo do poço e a vazão total de óleo no *topside*, ambas normalizadas no intervalo de 0 a 1, com base nas condições operacionais do poço e do sistema, enquanto a abertura da válvula *choke* e a injeção de gás (a_t) serão as variáveis manipuladas do processo.

Com o ambiente definido, o treinamento é estruturado em episódios de 10 horas, refletindo a dinâmica lenta do sistema de *gas lift*. Durante cada episódio, a razão gás-óleo (RGO) do poço varia, o que gera cenários perturbados que permitem explorar diferentes regimes de operação.

A recompensa (R_t) adotada neste trabalho foi definida como uma função de lucro, estabelecida pela diferença entre a receita obtida com a venda do óleo produzido e o custo associado ao consumo de gás injetado, multiplicada por uma penalização, conforme a restrição da Equação 4. Assim, a função de recompensa R_t é definida como:

$$\text{pen} = \begin{cases} (1 + 0.18 \cdot F), & \text{se } F < 0 \\ 1, & \text{caso contrário} \end{cases} \tag{5}$$

$$R_t = k \cdot (P_o\, w_{oil} - C_g\, w_{inj}) \cdot \text{pen} \tag{6}$$

onde P_o representa o preço do óleo, $w_{oil}(t)$ a vazão de óleo produzida, C_g o custo do gás e w_{inj} a vazão de gás injetado e k uma constante.

Dessa forma, o agente é incentivado a maximizar a função de recompensa apresentada na Equação 6, buscando um ponto de operação que aumente a vazão de óleo e, simultaneamente, reduza a quantidade de gás injetado, promovendo maior eficiência econômica do sistema, conforme ilustrado na Figura 2.

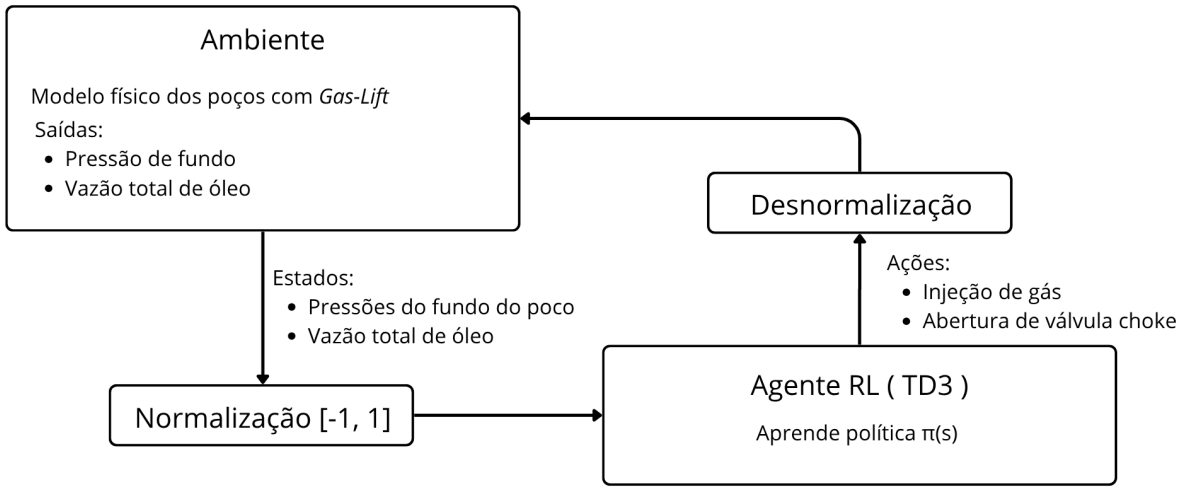

Figura 2: Diagrama da interação do agente RL com o ambiente.

4 Resultados e Discussão

Em uma primeira etapa, para analisar a resposta do agente de RL, foi realizada uma análise em regime estacionário da vazão de óleo em função da vazão de gás, com o objetivo de identificar o ponto ótimo de operação que maximize a produção de óleo no *topside*, em um regime estável e livre de golfadas. Considerando a Figura 3, observa-se que, para uma configuração com RGO igual a 0,06 e abertura máxima da válvula, o sistema permanece em regime de golfadas até aproximadamente 0,3 kg/s de injeção de gás. A partir desse ponto, emerge a curva característica do *gas lift* (relação entre vazão de óleo e injeção de gás), apresentando um platô em torno de 15,12 kg/s na vazão de óleo, indicando a região de operação próxima ao ótimo.

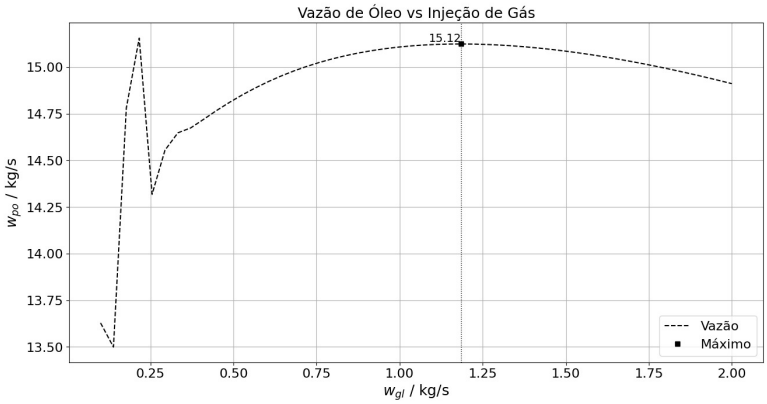

Figura 3: Análise do Ponto Ótimo

Ao realizar essa mesma configuração, partindo do mesmo valor de razão gásóleo, o agente de RL converge para uma ação caracterizada pela abertura total da válvula *choke* e por uma vazão de injeção de gás próxima de 1,26 kg/s (Figura 4). Esse comportamento é coerente com o objetivo de maximizar a produção de óleo, resultando em uma vazão final de aproximadamente 15,12 kg/s, correspondente à região ótima de operação do sistema. Observa-se ainda que, nos instantes iniciais da simulação, o sistema apresenta regime de golfadas, sendo posteriormente estabilizado após o acionamento do agente RL a partir das 2 horas.

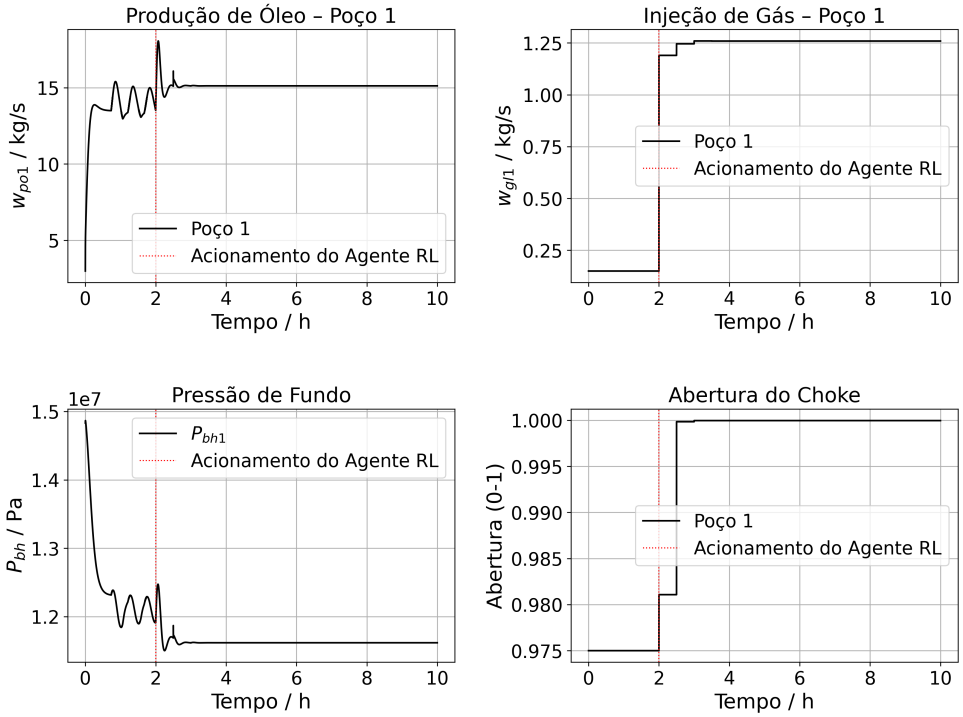

Figura 4: Desempenho do Agente RL

Para dar prosseguimento à análise de como esse agente reage a distúrbios e sua capacidade de evitar regime de golfadas, foram realizadas alterações na razão gás-óleo a cada 2 horas, fazendo o sistema começar a oscilar,

porém rapidamente convergindo para uma região sem golfadas.

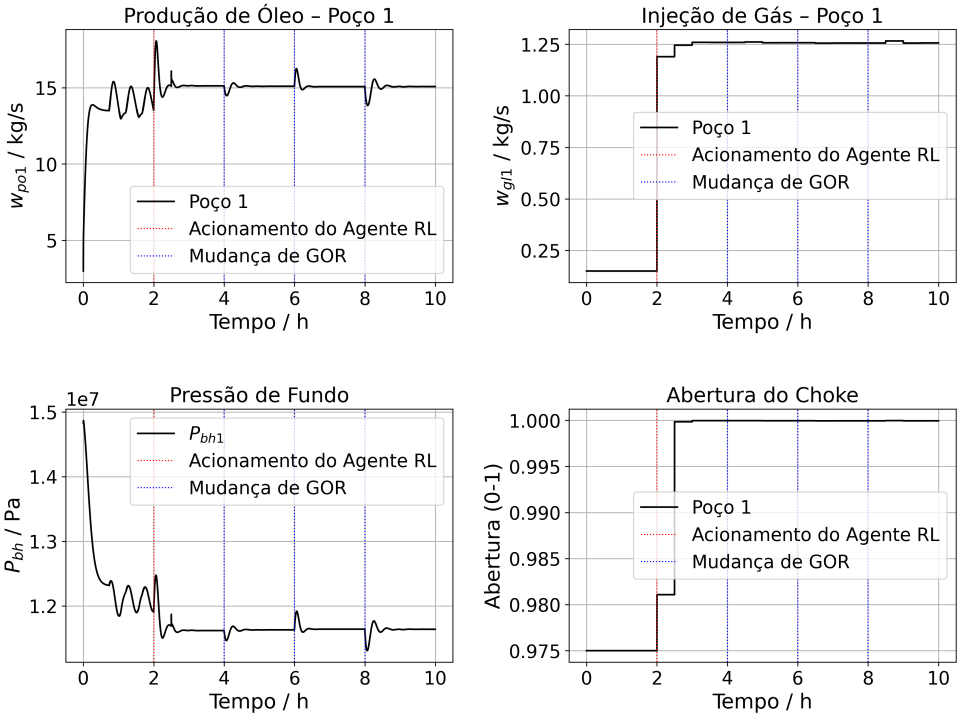

Figura 5: Desempenho do Agente RL com RGO variante

5 Conclusões

Este trabalho apresentou o desenvolvimento de um agente de aprendizado por reforço aplicado à otimização operacional de poços com *gas lift*, visando maximizar o lucro operacional, ao mesmo tempo em que respeita uma fronteira de operabilidade anti-golfadas. A política aprendida pelo agente foi obtida a partir da interação com o ambiente simulado, resultando em condições operacionais associadas à região economicamente ótima do sistema.

A política final convergiu para uma operação caracterizada pela abertura total da válvula *choke* e por uma vazão de gás injetado aproximadamente constante, indicando uma estratégia operacional economicamente eficiente. Além disso, os testes realizados sob variações na razão gásóleo demonstraram que o agente foi capaz de ajustar dinamicamente a injeção de gás, mantendo a produção de óleo próxima da região ótima de operação e evidenciando capacidade adaptativa frente a perturbações do processo.

Os resultados obtidos indicam que a abordagem proposta é promissora para problemas de otimização em sistemas de *gas lift*, que comportam regimes instáveis de escoamento. Entretanto, por se tratar de uma abordagem baseada em redes neurais artificiais, não há garantia formal de estabilidade ou segurança operacional. Dessa forma, para aplicações industriais, é recomendada a integração de estratégias com camadas adicionais de proteção, filtros de segurança ou arquiteturas híbridas envolvendo RTO (*Real-Time Optimization*) e controle avançado.

Agradecimentos: Os autores agradecem à Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), no âmbito PRH 41/UFBA, pelo suporte financeiro e apoio ao desenvolvimento deste trabalho.

Referências

Aranha Ribeiro, J. B., Vergara Dietrich, J. D., & Normey-Rico, J. E. (2024). Comparison of economic model predictive controllers for gas-lift optimization in offshore oil and gas rigs. *Computers Chemical Engineering*, *186*, 108685. https://doi.org/https://doi.org/10.1016/j.compchemeng.2024.108685

Cooksey, A., & Pool, M. (1995). Production Automation System For Gas Lift Wells. *SPE Production Operations Symposium*. https://doi.org/10.2118/29453-MS

de Rezende Faria, R., Capron, B. D. O., Secchi, A. R., & de Souza, M. B. J. (2024). Gas-Lift Optimization Using Physics-Informed Deep Reinforcement Learning. *Industrial & Engineering Chemistry Research*, *63*(32), 14199–14210. https://doi.org/10.1021/acs.iecr.3c04615

de Sousa Santos, L., de Souza, K. M. F., Bandeira, M. R., Ruiz Ahón, V. R., Peixoto, F. C., & Prata, D. M. (2018). Dynamic optimization of a continuous gas lift process using a mesh refining sequential method. *Journal of Petroleum Science and Engineering*, *165*, 161–170. https://doi.org/https://doi.org/10.1016/j.petrol.2018.02.019

Eikrem, G. O., Imsland, L. S., & Foss, B. A. (2004). Stabilization of Gas Lifted Wells Based on State Estimation. *IFAC Proceedings Volumes*, *37*, 323–328. https://api.semanticscholar.org/CorpusID:14680019

Fujimoto, S., Hoof, H., & Meger, D. (2018). Addressing Function Approximation Error in Actor-Critic Methods. *Proceedings of the International Conference on Machine Learning (ICML)*, 1587–1596.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2019). Continuous control with deep reinforcement learning. https://arxiv.org/abs/1509.02971

SOUZA, N. D. D., SOUZA, L. S. d., FONTES, R. M., MEIRA, R. L., & MARTINS, M. A. F. (2025). Análise de Operabilidade em Sistema de Gas-Lift Sujeito a Casing Heading [Acesso em: 04 maio 2026]. *Anais do Congresso Brasileiro de Engenharia Química e Encontro Brasileiro sobre o Ensino em Engenharia Química*, *5*. https://proceedings.science/cobeq/cobeq-2025/trabalhos/analise-de-operabilidade-em-sistema-de-gas-lift-sujeito-a-casing-heading?lang=pt-br