Journal of Oil and Gas Technology 石油天然气学报, 2025, 47(2), 283-293
Published Online June 2025 in Hans. https://www.hanspub.org/journal/jogt
https://doi.org/10.12677/jogt.2025.472032

Hans汉斯

基于近端策略优化算法多段压裂水平井CO_2吞吐注采优化

李荣涛*，曹小朋，张　东，李宗阳，王传飞，韩凤蕊，郭　祥

中国石化胜利油田分公司勘探开发研究院，山东　东营

收稿日期：2025年4月11日；录用日期：2025年6月16日；发布日期：2025年6月27日

摘　要

CO_2吞吐是致密油藏多段压裂水平井衰竭弹性开发后续提高采收率重要接替手段，吞吐注采优化具有成本低、易实现和效果明显的优点。目前吞吐注采优化方法存在不足，未充分考虑不同吞吐轮次和吞闷吐不同阶段间的干扰，本文建立了基于近端策略优化算法多段压裂水平井CO_2吞吐注采优化新方法，以净现值为优化目标，吞吐注采参数为优化变量，新方法实现了不同吞吐轮次变注采速度和变注采时长的动态注采优化，充分考虑了吞吐各阶段之间的干扰。实例Y区块CO_2吞吐注采优化结果表明：最优方案通过降低返排速度延长返排时间，充分提高返排阶段动用程度，提高了注入CO_2吞吐效率，减少吞吐轮次大幅降低注气成本，能够获得最优经济效益，为现场实际注采优化提供指导。

关键词

近端策略优化，多段压裂水平井，CO_2吞吐，注采优化，深度强化学习

CO_2 Huff and Puff Rates Optimization of Multi-Stage Fracturing Horizontal Wells Based on Proximal Policy Optimization Algorithm

Rongtao Li*, Xiaopeng Cao, Dong Zhang, Zongyang Li, Chuanfei Wang, Fengrui Han, Xiang Guo

Exploration and Development Research Institute, Sinopec Shengli Oilfield Branch, Dongying Shandong

Received: Apr. 11th, 2025; accepted: Jun. 16th, 2025; published: Jun. 27th, 2025

*通讯作者。

文章引用: 李荣涛, 曹小朋, 张东, 李宗阳, 王传飞, 韩凤蕊, 郭祥. 基于近端策略优化算法多段压裂水平井 CO_2 吞吐注采优化[J]. 石油天然气学报, 2025, 47(2): 283-293. DOI: 10.12677/jogt.2025.472032

Abstract

CO_2 huff and puff is an important replacement method for the subsequent improvement of oil recovery in the elastic development of multi-stage fracturing horizontal wells in tight reservoirs. The rates optimization of huff and puff injection and production has the advantages of low cost, easy implementation, and obvious effects. At present, the rates optimization method of huff and puff injection and production is insufficient, and the interference between different huff and puff cycles and different stages is not fully considered. A novel CO_2 huff and puff injection and production rates optimization method for multi-stage fractured horizontal wells based on the proximal policy optimization algorithm has been proposed. With the net present value as the optimization goal and huff and puff injection and production rates parameters as the optimization variables, the new method realizes dynamic injection and production rates optimization with different huff and puff cycles and variable injection and production speed and variable injection and production duration, and considers the interference between various stages. The injection and production rates optimization results of Y block CO_2 huff and puff indicate that the optimal project extends the backflow time by reducing the backflow speed, fully improving the utilization degree of the backflow stage. At the same time, improving the efficiency of CO_2 injection and reducing the number of cycles can significantly reduce gas injection costs, achieving optimal economic benefits and providing guidance for on-site actual CO_2 huff and puff injection and production rates optimization.

Keywords

Proximal Policy Optimization Algorithm, Multi-Stage Fracturing Horizontal Wells, CO_2 Huff and Puff, Rates Optimization, Deep Reinforcement Learning

Copyright © 2025 by author(s) and Hans Publishers 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

1. 引言

致密油藏因储层渗透率极低，需人工压裂进行储层改造，大量人工裂缝提供油气渗流高速通道，使得多段压裂水平井(MFHW)初期日产油量高达上百吨[1] [2]，但致密储层基质中原油动用程度低，造成弹性衰竭开发初期递减率高达 50%以上和采收率 10%左右[3]。连续气驱由于储层致密，注采井间难建立有效驱替，CO_2 吞吐成为提高采收率重要接替手段，补充地层能量且置换储层基质中原油[4] [5]。注采参数对 CO_2 吞吐效果影响明显，通过注采优化提高采收率具有成本低、易实施和效果好的优点。

目前，多段压裂水平井 CO_2 吞吐驱油机理实验研究较多[6]-[8]，矿场注采优化研究较少且存在不足[9] [10]。CO_2 吞吐不同轮次注气速度和返排速度保持不变，未能根据吞吐开发历程进行动态调整[11]-[13]。未对吞、闷和吐三个阶段生产时长进行优化，不同吞吐轮次的生产状况存在较大差异，且吞、闷和吐不同阶段之间存在干扰[14]，现有注采优化存在明显不足。因此，为实现不同吞吐轮次最优注采速度和注采时长的动态调整，本文建立了基于近端策略优化(PPO)算法的多段压裂水平井 CO_2 吞吐注采动态优化方法，弥补了现有 CO_2 吞吐注采优化的不足。

本文采用的深度强化学习 PPO 算法是一种无梯度优化方法，克服了有梯度优化算法获取优化目标对优化变量偏导数难实现的不足，能够与任意油藏数值模拟软件相结合，在实际应用中更易实现且应用范围广[15]。此外，PPO 算法基于马尔科夫链，将整个注采优化阶段划分为若干个注采调整时间步，能够对

每个注采调整时间步的注采参数进行动态调整，具有很强的注采动态优化优越性[16]。目前，深度强化学习在注采优化方面已有应用。经典离线策略强化学习 Q-learning 方法，已用于优化油藏概念模型一注一采两口井的注水速度[17]。PPO 算法已应用于优化井数、井位和钻井顺序等[18] [19]，其状态空间由流体饱和度图和压力分布场图组成。多智能体深度确定性策略梯度(MADDPG)算法用于多井 CO_2 连续驱注采优化，很好地考虑多井井间干扰，弥补了单智能体深度强化学习算法的不足[20]。

本文首先论述 PPO 算法原理和易收敛特性，适用于解决 CO_2 吞吐多轮次大幅度变化的注采参数优化问题。结合 PPO 算法原理和多段压裂水平井 CO_2 吞吐注采优化物理问题，明确了优化目标、优化决策和约束条件，建立了吞吐注采优化新方法，并应用于具体实例 Y 区块，论证了注采新方法的有效性，且深入分析注采优化增油机制，所得规律为矿场注采优化提供指导。

2. PPO 算法原理

近端策略优化方法由置信区间策略优化方法(TRPO)发展完善而来，TRPO 和 PPO 均属于基于策略的深度强化学习算法，成功解决了策略更新过程中确定合理学习步长的难题，因而在收敛性和稳定性方面表现出色，PPO 通常被谷歌 DeepMind 团队作为处理优化问题的首选算法。在约束新旧策略差别方面，PPO 比 TRPO 更易实现且求解效率更高[21]。

经过复杂的理论推导，TRPO 实现了策略更新过程中的单调提升，确保每次更新都带来进步。TRPO 通过最大化新旧策略之间的差值，以保证新策略在每次更新后均优于旧策略[22]。新旧策略的优劣利用价值期望进行量化表征。

$$\eta(\pi)=E_{s_0,a_0,s_1,a_1,\cdots}\left(\sum_{t=0}^{\infty}\gamma^t r(s_t)\right) \tag{1}$$

$$A_\pi(s_t,a_t)=Q_\pi(s_t,a_t)-V_\pi(s_t) \tag{2}$$

$$\begin{aligned}
&\sum_{\tau\sim\pi'}\left(\sum_{t=0}^{\infty}\gamma^t A_\pi(s_t,a_t)\right)\\
&=\sum_{\tau\sim\pi'}\left(\sum_{t=0}^{\infty}\gamma^t\left(Q_\pi(s_t,a_t)-V_\pi(s_t)\right)\right)\\
&=\sum_{\tau\sim\pi'}\left(\sum_{t=0}^{\infty}\gamma^t\left(r_{t+1}+\gamma V_\pi(S_{t+1})-V_\pi(s_t)\right)\right)\\
&=\eta(\pi')+E_{\tau\sim\pi'}\left(\sum_{t=0}^{\infty}\gamma^{t+1}V_\pi(S_{t+1})-\sum_{t=0}^{\infty}\gamma^t V_\pi(s_t)\right)\\
&=\eta(\pi')+E_{\tau\sim\pi'}\left(\sum_{t=1}^{\infty}\gamma^t V_\pi(S_t)-\sum_{t=0}^{\infty}\gamma^t V_\pi(s_t)\right)\\
&=\eta(\pi')-E_{\tau\sim\pi'}\left(V_\pi(s_0)\right)\\
&=\eta(\pi')-\eta(\pi)
\end{aligned} \tag{3}$$

其中，τ 为由状态概率、转换概率和策略决定的采样轨迹。

$$\begin{aligned}
J^\theta(\theta)&=\eta(\pi')-\eta(\pi)=\sum_{\tau\sim\pi'}\left(\sum_{t=0}^{\infty}\gamma^t A_\pi(s,a)\right)\\
&=\sum_{t=0}^{\infty}\sum_{s}P(s_t=s|\pi')\sum_{a}\pi'(a|s)\gamma^t A_\pi(s,a)\\
&=\sum_{s}\sum_{t=0}^{\infty}\gamma^t P(s_t=s|\pi')\sum_{a}\pi'(a|s)A_\pi(s,a)
\end{aligned}$$

$$= \sum_{s} \rho_{\pi'}(s) \sum_{a} \pi'(a|s) A_{\pi}(s,a) \tag{4}$$

其中，$\rho_{\pi}(s)$表示任意时间步状态为 s 概率和。

为了表征 TRPO 策略更新单调提升的特征，引入了优势函数，为给定状态下某一动作状态对价值与所有可能动作状态对价值平均值之间的差值见公式(2)。经推导得到优势函数与新旧策略差值之间的关系见公式(3)，发现当能确保新旧策略差值即优势函数为非负时，则能保证新策略优于旧策略性能单调提升。因此，TRPO 算法优化目标转化为最大化优势函数见公式(4)。

为了大幅提高样本利用率，TRPO 依据重要度采样定理，将在线学习策略转化为离线学习，使得旧策略生成的样本可被重复利用。经过重要度采样的转化，累积奖励期望表达式见公式(5)。通过引入重要度权重 $p(x)/q(x)$ 修正 $f(x)$，通过旧策略 $q(x)$ 分布，可计算新策略 $p(x)$ 分布的累积奖励期望，但需要确保新旧策略之间的动作概率分布相近。经过重要度采样的进一步转化，TRPO 优化目标如公式(6)所示。

$$\begin{aligned} E_{x\sim p}(f(x)) &= \int f(x)\,p(x)\,dx = \int f(x)\frac{p(x)}{q(x)}q(x)\,dx \\ &= E_{x\sim q}\left(f(x)\frac{p(x)}{q(x)}\right) \end{aligned} \tag{5}$$

$$J^{\theta'}(\theta) = E_{(s,a)\sim\pi_{\theta'}}\left(A_{\pi_{\theta'}}(s,a)\frac{\pi_{\theta}(s,a)}{\pi_{\theta'}(s,a)}\right) \tag{6}$$

TRPO 算法通过确保新旧策略期望差值非负，实现了策略更新的单调提升特性。将优化目标设定为新旧策略期望差值的最大化，为便于计算这一差值通过优势函数表示。TRPO 算法引入了重要度采样定理，以进一步提高样本利用率。通过使用旧策略分布替代新策略分布，将策略更新过程由在线学习转化为离线学习，重复采样历史旧策略产生的样本，但新旧策略分布差异不能过大，需遵循一定的约束条件。TRPO 算法与 PPO 算法在约束条件设置方面存在一些差异，PPO 算法则在此方面做出了重大改进，计算效率更高且更易于实施。

TRPO 算法通过 KL 散度约束新旧策略之间的差异程度，KL 散度通过利用两个分布之间的距离量化差异程度。TRPO 利用共轭梯度算法需要计算二阶偏导数进行求解，导致计算量大耗时过长。

$$\text{maximize } E_{(s,a)\sim\pi_{\theta'}}\left(A_{\pi_{\theta'}}(s,a)\frac{\pi_{\theta}(s,a)}{\pi_{\theta'}(s,a)}\right) \tag{7}$$

$$D_{\mathrm{KL}}^{\rho_{\pi_{\theta'}}}(\pi_{\theta'},\pi_{\theta}) \le \delta \tag{8}$$

PPO 算法对 TRPO 约束新旧策略差异方面进行改进，PPO 通过 Clip 函数对新旧策略重要度比值 $\pi_{\theta}(s,a)/\pi_{\theta'}(s,a)$ 进行截断处理，将其限制在一个特定范围内 $[1-\varepsilon,1+\varepsilon]$，以确保每次梯度更新波动幅度合理见公式(9)。同时，PPO 算法使用了最小值函数，以保证表现水平都能够达到优异。PPO 算法使用梯度上升法求解模型，需要计算一阶偏导数，求解难度降低，提高了计算效率。

$$\begin{aligned} J^{\theta'}(\theta) &= E_{(s,a)\sim\pi_{\theta'}}\left(\min\left(A_{\pi_{\theta'}}(s,a)\frac{\pi_{\theta}(s,a)}{\pi_{\theta'}(s,a)}, clip\left(A_{\pi_{\theta'}}(s,a)\left(\frac{\pi_{\theta}(s,a)}{\pi_{\theta'}(s,a)}\right)\right)\right)\right) \\ &\approx \sum_{(s,a)}\min\left(A_{\pi_{\theta'}}(s,a)\frac{\pi_{\theta}(s,a)}{\pi_{\theta'}(s,a)}, clip\left(A_{\pi_{\theta'}}(s,a)\left(\frac{\pi_{\theta}(s,a)}{\pi_{\theta'}(s,a)}\right)\right)\right) \end{aligned} \tag{9}$$

PPO 算法能够确定策略更新合理学习步长，保证新策略优于旧策略性能单调提升，与其他深度强化学习算法相比，具有收敛性强和稳定性高的优点。CO_2 吞吐注采参数不同轮次周期性强，吞、闷和吐不同生产阶段注采参数变化幅度大，对注采优化算法性能要求高。注采优化物理问题需与优化方法相匹配，因此，选用收敛性强和稳定性高的 PPO 算法作为多段压裂水平井 CO_2 吞吐注采优化方法。

PPO 算法能够通过限定新旧策略间的差异，确保策略更新幅度处于可控区间，进而实现策略性能的单调提升，具备卓越的稳定性和收敛性。CO_2 吞吐注采参数随不同轮次呈现周期性变化，且吞、闷、吐各阶段注采参数波动幅度大，对优化算法性能要求严苛。注采优化物理问题需与优化方法相匹配，因此，选用收敛性强和稳定性高的 PPO 算法作为多段压裂水平井 CO_2 吞吐注采优化方法。对于不同区块、不同油藏类型的注采参数优化需求，PPO 算法依靠其强稳定性和强收敛性，均能依据油藏地质特征与开发状况，通过科学构建强化学习环境及合理制定奖励函数，精准优化 CO_2 吞吐注采参数，从而显著提升开发效果，为油气田高效开发提供有力的技术支撑。

3. 注采优化方法

基于 PPO 的多段压裂水平井 CO_2 吞吐注采优化为典型的最优化问题，包含优化目标、优化决策和约束条件三个要素。注采优化以净现值为目标，追求经济效益最大化。优化决策为吞、闷和吐不同生产阶段的注采参数，包括注气速度、注气时长、闷井时间、返排速度和返排时间。状态空间变量为吞吐井生产数据，用以描述吞吐注采环境主要特征，为智能体优化决策提供信息。所求注采参数最优解需满足一系列约束条件限制。

3.1. 优化目标

综合考虑注采优化物理问题和 PPO 优化算法原理，确定 CO_2 吞吐注采优化目标。若以采收率为注采优化目标，则智能体为追求采收率最大化，大幅增加累注气量，导致换油率低经济效益差。若以换油率为注采优化目标，换油率随生产时间增大而降低，增加了 PPO 算法收敛难度。因此，选择以净现值为注采优化目标见公式(10)，兼顾采收率和注气成本两方面，追求经济效益最大化。

$$NPV=\sum_{n=1}^{N}\frac{P_oQ_o^{\Delta t^n}-C_{CO_2-INJ}Q_{CO_2-INJ}^{\Delta t^n}-C_{CO_2-PRO}Q_{CO_2-PRO}^{\Delta t^n}}{\left(1+b\right)^{t^n/365}} \tag{10}$$

其中，P_o 为原油价格，元/吨；C_{CO_2-INJ} 和 C_{CO_2-PRO} 分别为注入 CO_2 价格和产出 CO_2 处理费用，元/吨；$Q_o^{\Delta t^n}$、$Q_{CO_2-INJ}^{\Delta t^n}$ 和 $Q_{CO_2-PRO}^{\Delta t^n}$ 分别为第 n 个注采调整时间步的累产油量、CO_2 累注入量和 CO_2 累产量，吨；N 为注采优化调整时间步总数；t^n 为第 n 注采调整时间步的时长，天；b 为年基准利率，%。

3.2. 优化决策

在深度强化学习 PPO 算法框架下，需要明确动作空间变量和状态空间变量。动作空间变量表征物理优化问题的决策变量，为 CO_2 吞吐不同生产阶段的注采速度和注采时长，具体包括注气速度、注气时间、闷井时间、返排速度和返排时间。为了避免注采优化变量变化幅度过大，造成现场施工困难和油藏模拟计算难收敛，限定相邻调整时间步注采参数最大变化幅度在正负 20%以内。

状态空间变量描述当前时间步注采环境状态特征，为智能体选择最优注采动作提供重要决策信息。选取 CO_2 吞吐井生产数据作为状态空间变量，包括生产时间、生产气油比、井底流压、日注气量、日产气量、日产油量、累注气量、累产气量和累产油量，共包括 9 个井生产数据参数。本文未采用饱和度场图和压力场图作为状态空间变量，因为场图数据精度受地质模型不确定性影响明显，难以保证精度且获取难度较大。井生产数据现场获取容易且精度高。此外，场图数据训练神经网络所需时间较长，计算效

率较低且耗时较长。井生产数据样本量较少，神经网络训练耗时短效率高。

3.3. 约束条件

多段压裂水平井 CO_2 吞吐注采优化最优解需满足约束条件限制。注入阶段井底流压上限低于岩石破裂压力的 90%。返排阶段井底流压下限设定为原油泡点压力附近，井底流压下限较低可获得更大生产压差，能够更充分地利用地层能量，提高储层动用程度。但井底流压下限过低，压敏效应造成储存伤害严重。因此，返排阶段井底流压下限需设置合理，兼顾提高生产压差和压敏储层伤害两方面。

3.4. 优化流程

通过智能体与注采环境之间不断交互，得到多段压裂水平井 CO_2 吞吐注采参数最优解。CO_2 吞吐井作为智能体，根据当前时间步的注采状态空间变量提供信息，动作选择神经网络遵循策略选择最优注采参数，选择使累积奖励净现值最大化的动作，动作选择神经网络朝着使累积净现值最大化的方向调整权重。每个注采调整时间步智能体将动作变量注采参数传递到油藏数值模拟器，经过油藏数值模拟计算获得吞吐井生产数据，进而计算本注采调整时间步的即时奖励阶段净现值，并构成下一个注采调整时间步的状态空间变量。这些数据作为样本存储在经验池中，用于训练后续回合神经网络。后续智能体与环境交互回合，目标动作选择网络在历史经验池中寻找使累积奖励最大化的最优注采参数。在线动作选择网络朝着使累积奖励最大化的方向调整神经网络权重。随着回合次数的增加，优化目标累积净现值逐渐收敛到最大值，此时获得最优注采参数，注采优化算法流程如图 1 所示。

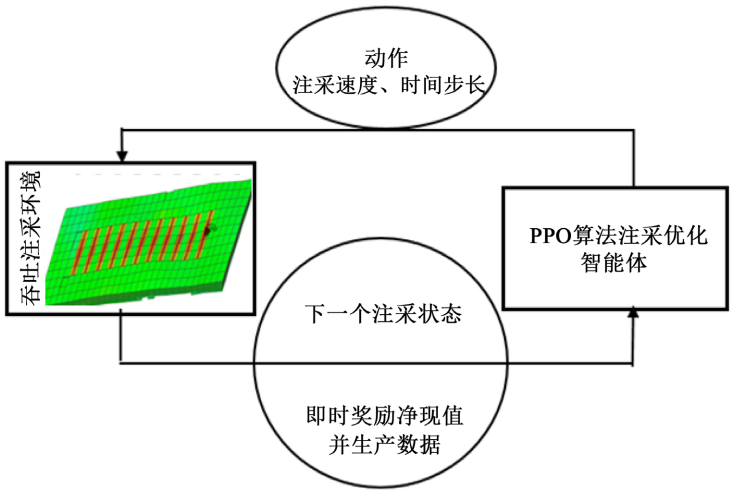

Figure 1. Optimization process of MFHW-CO_2 huff and puff rates based on PPO
图 1. 基于 PPO 的 MFHW-CO_2 吞吐注采优化流程

4. 实例应用

以 Y 区块为实例模型，应用基于 PPO 的多段压裂水平井 CO_2 吞吐注采优化新方法，得到最优注采参数。分析吞、闷和吐不同生产阶段的注采优化增油机制，为矿场实际生产提供指导。Y 区块主要受浊积水道控制，主要岩性为粉细和细粒长石砂岩。油层埋深 2100 m，油层平均厚度 23.3 m。平均渗透率 0.41 mD，平均孔隙度 0.11。油藏温度 70.6℃，原始地层压力 15.1 MPa 如图 2 所示。储层基质天然裂缝不发育，利用局部网格加密模拟了人工压裂裂缝，人工主裂缝模拟缝宽的确定，依据裂缝导流能力进行等效处理。

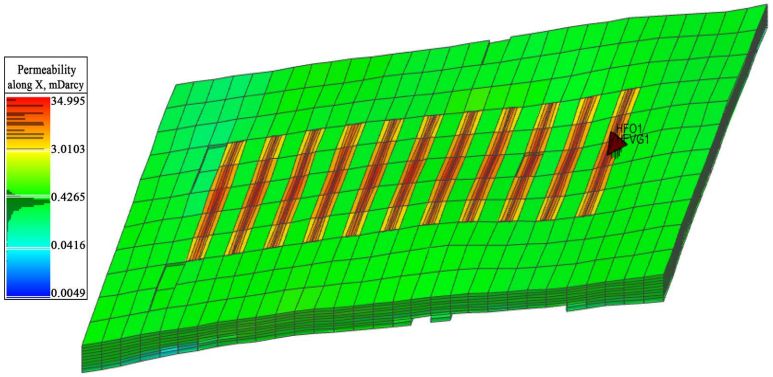

Figure 2. MFHW-CO_2 huff and puff calculation model of Y block
图 2. Y 块 MFHW-CO_2 吞吐计算模型

优化目标净现值的计算经济参数取值如下：原油价格 2818 元/吨，注入 CO_2 价格 550 元/吨，产出 CO_2 处理价格 30 元/吨，年基准利率 8%。CO_2 吞吐井生产数据作为状态空间变量，CO_2 吞吐注采速度和注采时长作为动作空间变量。PPO 算法基于 Actor-Critic 算法框架，Actor 动作选择神经网络有 3 层，中间层 64 个神经元，输入层神经元数量等于状态空间维度 9，输出层神经元数量等于动作空间维度 2。Critic 价值评价神经网络用于评估注采状态和动作变量的价值，包含 3 层神经网络，中间层 64 个神经元，输入层神经元数量等于智能体状态空间维度 9，输出层神经元数量等于价值维度 1。

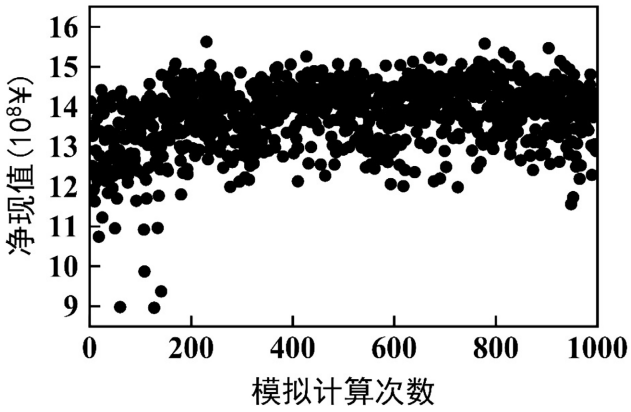

Figure 3. Convergence change of Y block rates optimization target NPV
图 3. Y 区块注采优化目标净现值收敛变化

多段压裂水平井 CO_2 吞吐实例注采优化目标累积净现值收敛变化如图 3 所示。随着模拟回合次数的增加，优化目标累积净现值在第 400 个回合收敛到最大值，并保持在最大值附近波动，直到第 1000 个回合结束。为了论证注采优化方法的有效性，将优于 95%的优化方案与基准方案进行比较。基准方案注采参数保持初值不发生变化，P95 最优方案注采参数经过优化动态变化。Y 实例区块基准方案和 P95 最优方案的累积净现值分别为 14.01×10^8¥和 14.92×10^8¥，P95 最优方案净现值比基准方案高出 6.50%，这证明了注采优化新方法的有效性。

首先，分析 P95 最优方案比基准方案净现值更高的原因。通过对比分析累产数据和累注数据如图 4 所示，发现 P95 最优方案的累产油量略低于基准方案，说明两方案由累产油带来的经济收益相差很小。累注气量和累产气量明显低于基准方案，优化方案由累注气产生的经济投入成本更低。因此，P95 最优方

案经济效益更高，获得了更高的净现值。

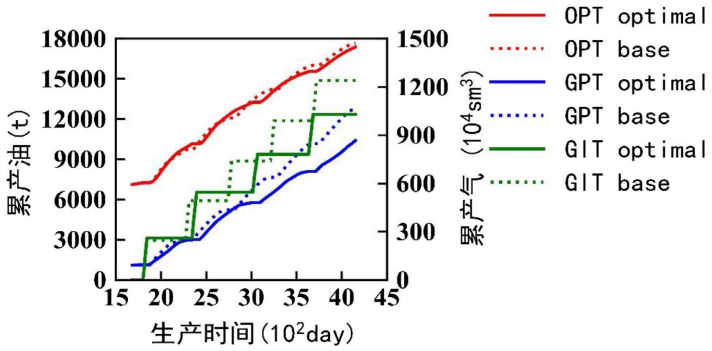

Figure 4. Cumulative production data comparison between base case and optimal case
图 4. 基准方案和最优方案累产数据对比

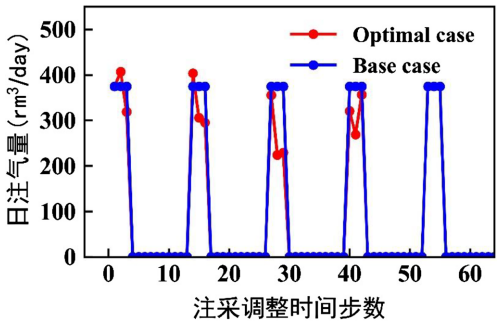

Figure 5. Gas injection rate comparison between base case and optimal case
图 5. 基准方案和最优方案日注气量对比

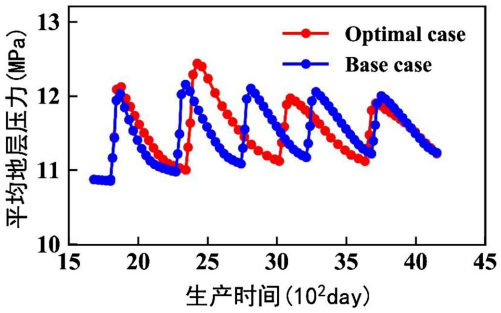

Figure 6. Average formation pressure between base case and optimal case
图 6. 基准方案和最优方案平均地层压力对比

然后，分析优化方案与基准方案累产油相差小的原因，通过对比多段压裂水平井 CO_2 吞吐不同生产阶段差异。在注入阶段，对比基准方案和优化方案的日注气量发现，整体上 P95 优化方案的日注气量明显低于基准方案，仅在第一个和第二个吞吐轮次初期略高于基准方案如图 5 所示。基准方案平均地层压力略高于优化方案，仅第一个和第二个吞吐轮次优化方案低于优化方案，其他吞吐轮次高于优化方案如图 6 所示。P95 优化方案比基准方案减少一个吞吐轮次，因此，优化方案累注气量明显更少，对应注气成本更低，考虑到累产油相差小，其换油率更高经济效益更好。但平均地层压力更低，为返排阶段提供的地层能量更少。井底流压下限设定在原油泡点压力附近，将由压敏造成的储层伤害控制在一定范围内。

综合来看，优化方案累注气量更少，注气成本更低经济效益更高，压敏造成储层伤害不利影响较低。

闷井阶段 CO_2 通过分子扩散与基质中的原油发生作用。对比基准方案和优化方案闷井时间相差很小如图 7 所示。另一方面，由于 Y 区块致密储层基质渗透率极低且天然裂缝不发育，CO_2 分子扩散很难进入到基质深处与更多原油发生作用。因此，闷井时间对 Y 区块吞吐累产油影响很小可忽略。

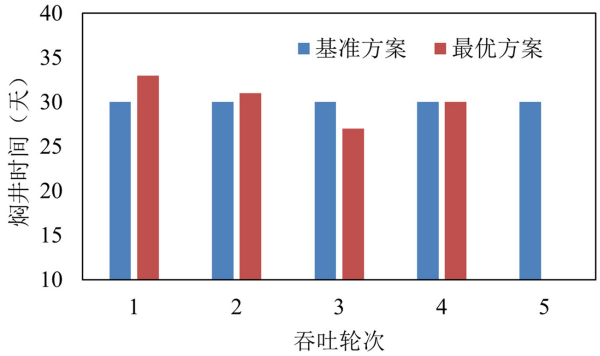

Figure 7. Soaking time comparison between base case and optimal case
图 7. 基准方案和最优方案闷井时间对比

对比基准方案和 P95 优化方案返排生产阶段，发现优化方案的返排速度明显低于基准方案，而返排时间步长明显高于基准方案如图 8 和图 9 所示。这表明优化方案通过降低返排速度和延长返排时间，提高了返排阶段的动用程度，使得最优方案在累注气量明显减少的情况下，累产油量仍能与基准方案基本持平。

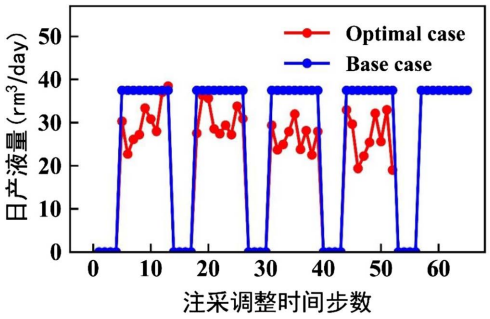

Figure 8. Liquid production rate comparison between base case and optimal case
图 8. 基准方案和最优方案日产液量对比

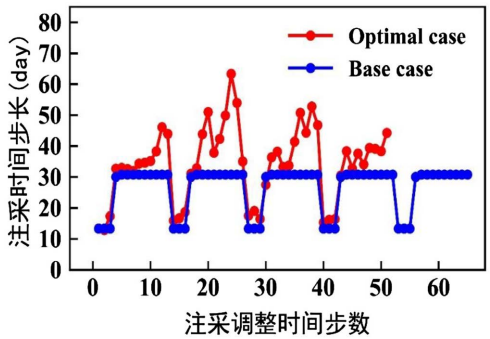

Figure 9. Time step comparison between base case and optimal case
图 9. 基准方案和最优方案注采时间步长对比

综上，实例 Y 区块多段压裂水平井 CO_2 吞吐注采优化方案与基准方案累产油基本持平，因为累注气量减少和压敏效应储层伤害带来的负面影响，与通过降低返排速度和延长返排时间，提高返排阶段动用程度带来的积极影响相互抵消。保持累产油量基本持平，注采优化方案减少了吞吐轮次，显著降低了累注气量，降低了注气成本，从而获得了更好的经济效益，因此其净现值高于基准方案。

5. 结论

(1) PPO 算法收敛性强和稳定性高，适用于解决 CO_2 吞吐周期性强和变化幅度大的注采优化问题。建立了基于 PPO 的多段压裂水平井 CO_2 吞吐注采优化方法，并应用于 Y 实例区块验证了新方法的有效性。

(2) 多段压裂水平井 CO_2 吞吐注采优化方案，通过降低返排产液速度和延长返排时间，提高返排阶段的动用程度，抵消掉累注气量降低和压敏储层伤害的负面影响，获得与基准方案基本持平的累产油量。优化方案累注气量更低，注气投入成本更低，因此，获得更高的累积净现值。

基金项目

区域二氧化碳捕集与封存关键技术研发与示范，十四五国家重点研发计划项目，项目编号：2022YFE0206800；

基于深度学习的 WAG-CO_2 注采优化方法研究，中国石化胜利油田分公司博士后项目，项目编号：YKB2406。

参考文献

[1] 张君峰, 毕海滨, 许浩, 等. 国外致密油勘探开发新进展及借鉴意义[J]. 石油学报, 2015, 36(2): 127-137.

[2] 贾承造, 邹才能, 李建忠, 等. 中国致密油评价标准主要类型基本特征及资源前景[J]. 石油学报, 2012, 33(3): 343-350.

[3] 宋俊强, 李晓山, 尤浩宇, 等. 玛湖砾岩油藏复模态结构下的压裂水平井压力分析[J]. 科学技术与工程, 2022, 22(19): 8295-8303.

[4] Ding, M., Gao, M., Wang, Y., Qu, Z. and Chen, X. (2019) Experimental Study on CO_2-EOR in Fractured Reservoirs: Influence of Fracture Density, Miscibility and Production Scheme. *Journal of Petroleum Science and Engineering*, **174**, 476-485. https://doi.org/10.1016/j.petrol.2018.11.039

[5] Zuloaga, P., Yu, W., Miao, J. and Sepehrnoori, K. (2017) Performance Evaluation of CO_2 Huff-N-Puff and Continuous CO_2 Injection in Tight Oil Reservoirs. *Energy*, **134**, 181-192. https://doi.org/10.1016/j.energy.2017.06.028

[6] 侯广. 致密油体积压裂水平井 CO_2 吞吐实践与认识[J]. 大庆石油地质与开发, 2018, 37(3): 163-167.

[7] 何应付, 赵淑霞, 刘学伟. 致密油藏多级压裂水平井 CO_2 吞吐机理[J]. 断块油气田, 2018, 25(6): 752-756.

[8] 杨正明, 刘学伟, 张仲宏, 等. 致密油藏分段压裂水平井注二氧化碳吞吐物理模拟[J]. 石油学报, 2015, 36(6): 724-729.

[9] Sun, J., Zou, A., Sotelo, E. and Schechter, D. (2016) Numerical Simulation of CO_2 Huff-N-Puff in Complex Fracture Networks of Unconventional Liquid Reservoirs. *Journal of Natural Gas Science and Engineering*, **31**, 481-492. https://doi.org/10.1016/j.jngse.2016.03.032

[10] Alfarge, D., Wei, M., Bai, B. and Almansour, A. (2017) Effect of Molecular-Diffusion Mechanisim on CO_2 Huff-N-Puff Process in Shale-Oil Reservoirs. *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*, Dammam, 24-27 April 2017, SPE-188003-MS. https://doi.org/10.2118/188003-ms

[11] Sanchez-Rivera, D., Mohanty, K. and Balhoff, M. (2015) Reservoir Simulation and Optimization of Huff-and-Puff Operations in the Bakken Shale. *Fuel*, **147**, 82-94. https://doi.org/10.1016/j.fuel.2014.12.062

[12] Yu, W., Zhang, Y., Varavei, A., Sepehrnoori, K., Zhang, T., Wu, K., *et al*. (2019) Compositional Simulation of CO_2 Huff 'n' Puff in Eagle Ford Tight Oil Reservoirs with CO_2 Molecular Diffusion, Nanopore Confinement, and Complex Natural Fractures. *SPE Reservoir Evaluation & Engineering*, **22**, 492-508. https://doi.org/10.2118/190325-pa

[13] Li, L., Zhang, Y. and Sheng, J.J. (2017) Effect of the Injection Pressure on Enhancing Oil Recovery in Shale Cores during the Co2 Huff-N-Puff Process When It Is above and Below the Minimum Miscibility Pressure. *Energy & Fuels*, **31**, 3856-3867. https://doi.org/10.1021/acs.energyfuels.7b00031

[14] Chen, C. and Gu, M. (2017) Investigation of Cyclic CO_2 Huff-and-Puff Recovery in Shale Oil Reservoirs Using Reservoir Simulation and Sensitivity Analysis. *Fuel*, **188**, 102-111. https://doi.org/10.1016/j.fuel.2016.10.006

[15] Sutton, R.S. and Barto, A.G. (2018) Reinforcement Learning: An Introduction. MIT Press.

[16] Peters, J. (2010) Policy Gradient Methods. *Scholarpedia*, **5**, 3698. https://doi.org/10.4249/scholarpedia.3698

[17] Talavera, A.L., Túpac, Y.J. and Vellasco, M.M. (2010) Controlling Oil Production in Smart Wells by MPC Strategy with Reinforcement Learning. *SPE Latin American and Caribbean Petroleum Engineering Conference*, Lima, 1-3 December 2010, Peru, SPE-139299-MS. https://doi.org/10.2118/139299-ms

[18] Nasir, Y. (2020) Deep Reinforcement Learning for Field Development Optimization. arXiv: 2008.12627.

[19] Nasir, Y., He, J., Hu, C., Tanaka, S., Wang, K. and Wen, X. (2021) Deep Reinforcement Learning for Constrained Field Development Optimization in Subsurface Two-Phase Flow. *Frontiers in Applied Mathematics and Statistics*, **7**, Article 689934. https://doi.org/10.3389/fams.2021.689934

[20] Rongtao, L., Liao, X., Wang, X., Zhang, Y., Mu, L., Dong, P., *et al*. (2022) A Multi-Agent Deep Reinforcement Learning Method for CO_2 Flooding Rates Optimization. *Energy Exploration & Exploitation*, **41**, 224-245. https://doi.org/10.1177/01445987221112235

[21] Schulman, J., Wolski, F,. Dhariwal, P., *et al*. (2017) Proximal Policy Optimization Algorithms. arXiv: 1707.06347.

[22] Schulman, J., Levine, S., Abbeel, P., *et al*. (2015) Trust Region Policy Optimization. arXiv: 1502.05477.