Closed-Loop CO_2 Storage Control With History-Based Reinforcement Learning and Latent Model-Based Adaptation

Sofianos Panagiotis Fotias[1] and **Vassilis Gaganis**[1,2,#,*]

[1] Institute of Geoenergy, Foundation for Research and Technology, Chania 73100, Greece

[2] School of Mining and Metallurgical Engineering, National Technical University of Athens, Athens 15773, Greece

Presenting author: Vassilis Gaganis, email: sfotias@metal.ntua.gr
* Corresponding author: Vassilis Gaganis, email: sfotias@metal.ntua.gr

ABSTRACT

Closed-loop control of geologic CO_2 storage must be robust to geological uncertainty while operating with limited monitoring, where dense subsurface states (e.g., pressure and saturation fields) are not available at deployment. We present a model-free reinforcement learning (RL) workflow that trains controllers on an ensemble of high-fidelity digital twins and evaluates them on a held-out reference simulator used strictly for inference. At each control step, the agent selects multi-well injection/production set-points and is rewarded for storage-oriented objectives while penalizing operational and pressure-related constraint violations, consistent with practical CCS control.

We organize the study around progressively reducing information and quantifying its effect on closed-loop performance. As an upper bound, we train an “oracle” Soft Actor-Critic (SAC) controller that conditions on both 3D reservoir state and well observations. In our runs, the oracle reaches a near-optimal return of ≈19.61 cumulative episode reward within 30 training epochs (where each epoch uses a fixed number of gradient updates, allowing fair comparison across variants). As a deployable lower bound, we train a wells-only SAC controller; it attains a best return of ≈18.82 but shows strong epoch-to-epoch variability with sharp regressions during training, reflecting the difficulty of controlling CCS using sparse well measurements alone.

To address partial observability using only sparse well measurements, we introduce a history-as-context attention architecture (GTrXL-like) within off-policy SAC: at each decision step, the policy and value networks receive a fixed window of past well observations rather than a single snapshot. This improves training stability and reaches the near-optimal regime (best return ≈19.65) within the 30-epoch budget, indicating that short-term operational history can compensate for missing 3D subsurface information.

We then introduce two mechanisms to understand and close the observability gap. First, curriculum masking progressively occludes the 3D state during training. This produces an “information-collapse” characterization: performance remains in the near-optimal band (here, returns 19.5) under partial masking, but regresses toward the wells-only regime when masking becomes complete (final best reward ≈18.95). Second, we propose an asymmetric knowledge-distillation scheme: during training on digital twins, the critic (teacher) retains access to privileged 3D state while the actor (student) is restricted to well-observation histories only; distillation aligns the student’s learning signal with the teacher’s value estimates. This yields a deployable controller that recovers near-oracle performance, reaching ≈19.62 by the end of training.

Overall, the progression provides practical guidance on observability requirements and mitigation strategies for CCS control under uncertainty.