A Comparative Analysis of Reinforcement Learning Methods for Stratospheric Balloon Station Keeping

Sean Corum and Jared Watrous

CS234 Reinforcement Learning Spring 2024 Final Project

Synopsis

Stratospheric balloons navigate by using ballasting to elevate or lower in a wind field to station keep near a target. Traditionally, these balloons have been piloted by ground based human operators or heuristic autopilot programs. Recently, reinforcement learning has been used to create AI autopilot programs. The state-of-the-art for this application domain includes the following RL algorithms: Deep Q-Network (DQN), Quantile Regression DQN (QR-DQN), and Soft Actor Critic (SAC). In this project, we sought to reproduce recent results in the literature within a common simulation environment. We also implement Proximal Policy Optimization (PPO), a novel use of this algorithm in the stratospheric balloon domain.
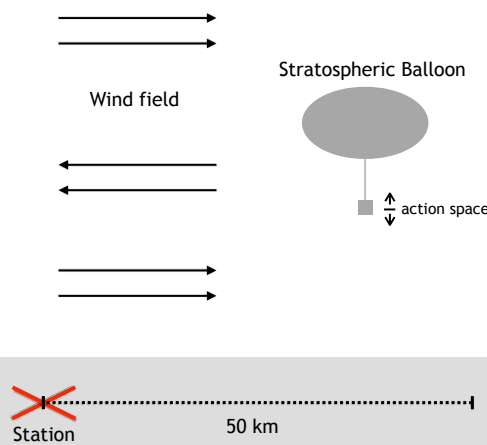
Algorithms
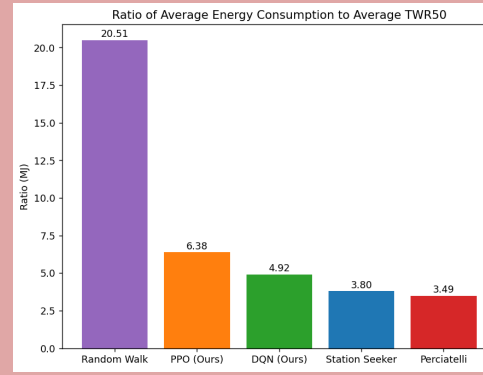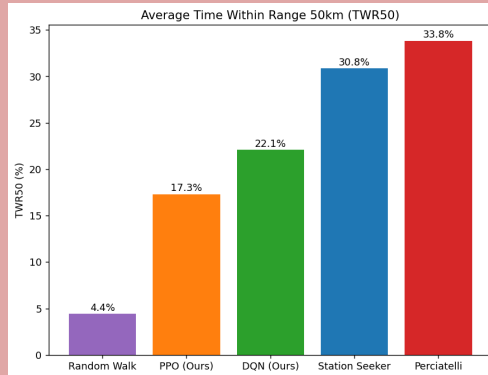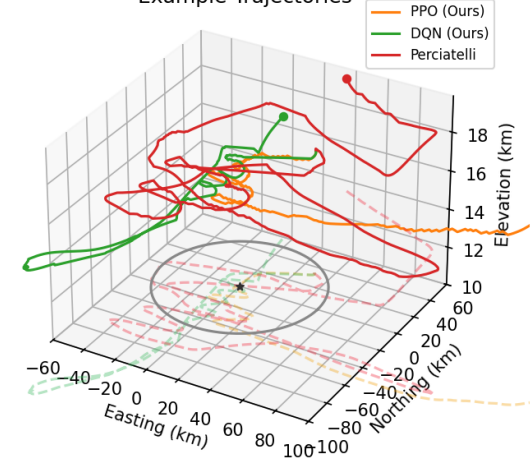
- PPO
- DQN
- SAC (future work)

Controls

- Random Walk
- StationSeeker (heuristic autopilot)
- Perciatelli (published QR-DQN model)

Station Keeping - moving up and down in the wind field to stay within 50 km of a station. Wind direction is a function of elevation. **TWR50** = % time within 50 km of station.

Wind field

Stratospheric Balloon

action space

Station — 50 km

Example Trajectories

Model Training

Average Time Within Range 50km (TWR50)

	TWR50
Random Walk	4.4%
PPO (Ours)	17.3%
DQN (Ours)	22.1%
Station Seeker	30.8%
Perciatelli	33.8%

Ratio of Average Energy Consumption to Average TWR50

	Ratio (MJ)
Random Walk	20.51
PPO (Ours)	6.38
DQN (Ours)	4.92
Station Seeker	3.80
Perciatelli	3.49

Takeaways

- Thus far we implemented PPO and DQN within the simulated environment, where they learned
- Have not been able to match the state-of-the-art
- Future work includes implementing SAC and further model optimization

References

- Bellemere et al. Autonomous navigation of stratospheric balloons using reinforcement learning. *Nature*, 588(7836):77-82, 2020.