CEORL

Cost of Energy Optimisation with Reinforcement Learning

Control Challenges Addressed

- Model-based control is sensitive to model quality.
- Sea-state control ('constant damping') is ineffective.
- Optimising power capture does not optimise LCOE.

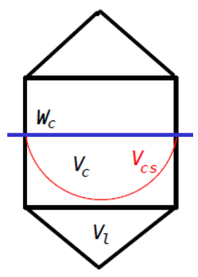
FastHeavingBuoy Environment

FastHeavingBuoy model:

- Simple m-s-d model in heave, PTO relative to ground.
- Bretschneider spectrum, H_s 2.75m, T_e 8.5s (different phases).
- No amplitude limitations in hydrodynamics.

Reward = Energy - Penalty

- Penalty represents cost of large swept volumes.
- Penalty proportional to amount by which the magnitude of (displacement - wave elevation) exceeds a threshold, β.

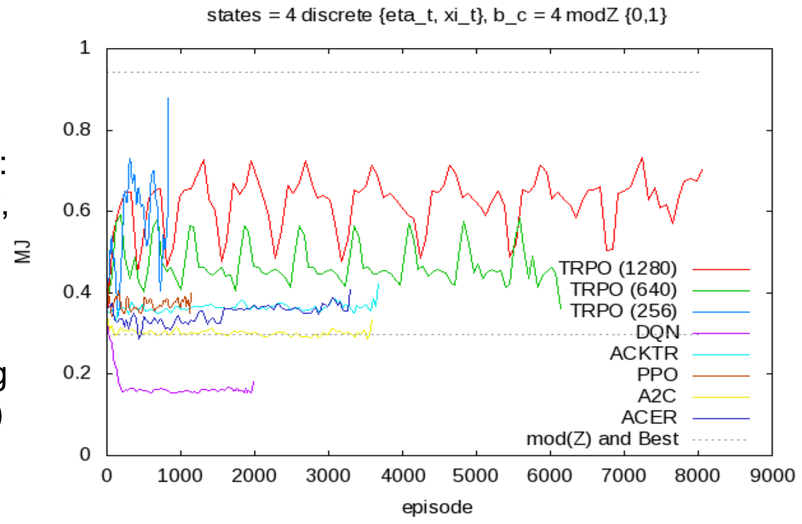


Algorithm investigations

states = 4 discrete {eta_t, xi_t}, b_c = 4 modZ {0,1}

Reinforcement Learning:

- 4 discrete states: position, velocity, wave elevation (η), d/dt of η.
- 2 discrete actions: damping = either 4mod(Z) or 0.



Model: radius 1m, draft and freeboard 1.5m, β = 1m.

Lower dashes: best constant damping control: mod(Z).

Upper dashes: best state-dependant control found by grid search.

TRPO and PPO identified as promising candidates:

- Yet did not converge on the best policy.

Contributors

1. MaxSim: Paul Stansell,
2. Power Enable Solutions Ltd: Richard Crozier,
3. Marine Systems Modelling: Joseph van t' Hoff,
4. The University of Edinburgh: David Forehand,
5. Wave Conundrums Consulting: Alexandra Price,
6. Aquaharmonics: Max Ginsburg and Alex Hagmüller,
7. Caelulum: Max Carcas,
8. Mocean Energy: Chris Retzler
9. CorPower Ocean: Jørgen Hals Todalshaug
10. David Pizer
11. Quoceant: Ross Henderson

FastHeavingBuoy – damping only

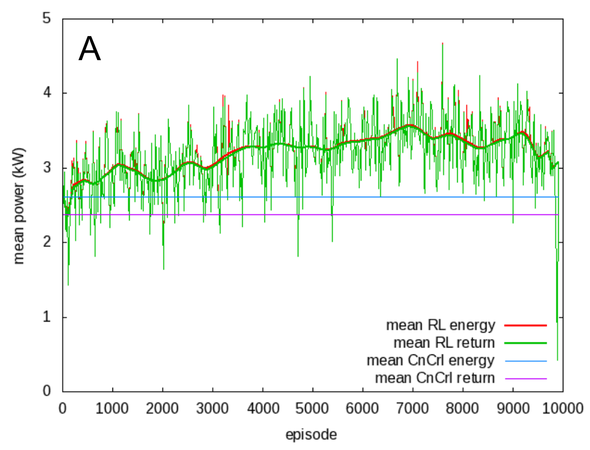
Model: radius 1m, draft and freeboard 1.5m, β = 1m.

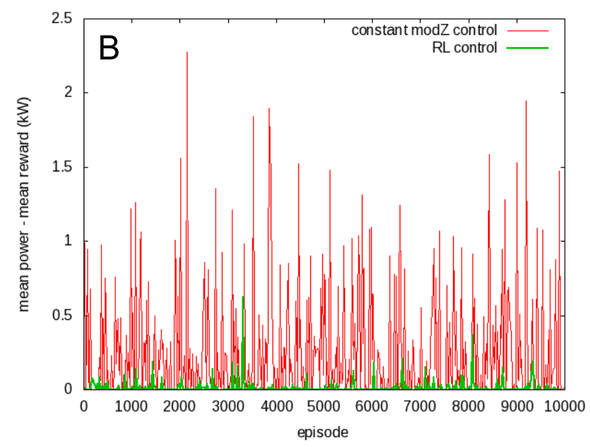
Reinforcement learning:

- 2D continuous state space: velocity and d/dt of η.
- 1D continuous action space: damping set between 0 and 4 mod(Z).
- Benchmark: continuous control found using grid search.
- Uses distributional C51 RL algorithm.

Plot A: more power capture

- Improvement in reward 40% - 45%; and power capture 30% - 35%.
- In RL policy, energy and reward are close together: WEC rarely enters potentially damaging regions of its operation.





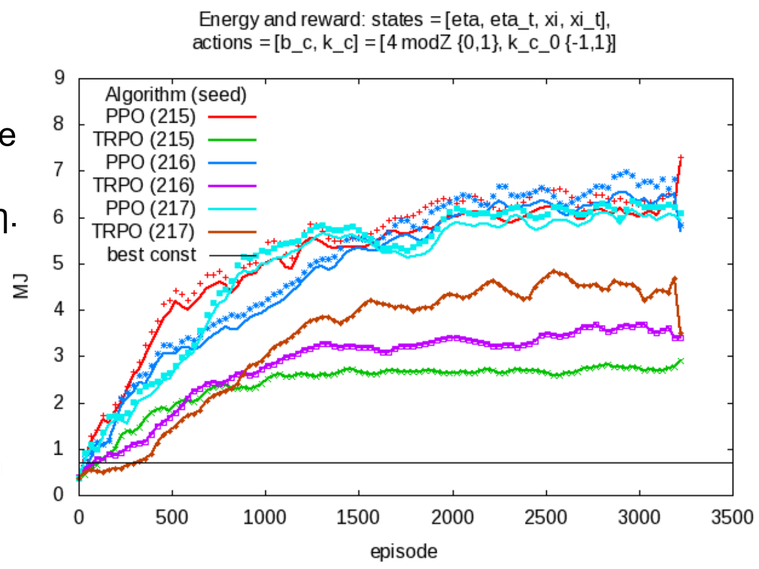
Plot B: fewer extreme motions

- Shows mean reward subtracted from mean power.
- Even though RL policies are producing more power, they are also reducing operation in potentially damaging operational regions.

FastHeavingBuoy–damping & spring

Reinforcement Learning:

- 4D continuous state space: position, velocity, η, d/dt of η.
- 2D continuous action space: PTO damping set between 0 and 4 mod(Z); and PTO spring set between ±5.3 times hydrostatic spring.



Model: radius 1m, draft and freeboard 3.5m, β = 3.25m.

Benchmark (0.7MJ): best constant control using damping and spring in ranges available to RL.

Results for most recent algorithm tests:

- For each colour, lines are rewards, symbols are energy.
- PTO damping and spring allowed in both RL and best constant control found with a grid search.
- PPO algorithm gives 800% improvement on best constant control.
- Note impact of starting seed on converged policy.

Contact:
paulstansell@gmail.com