Can Reinforcement Learning Support Policy Makers?

TL;DR – **YES!** Assessment models are essential for climate-related policy making, but they are also complicated and computationally demanding. Training RL agents to function as policy makers is a cheap and straightforward way to evaluate their emergent properties, and explore their solution space.

Théodore Wolf[1,2], Nantas Nardelli[1], John Shawe-Taylor[2], María Pérez Ortiz[2]

2023 NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

- Climate models are extremely complex, often intractable, dynamical models.
- **Integrated Assessment Models** (IAMs) are climate models that also integrate human behaviour (e.g. econometrics, demographics). They are higher fidelity, but even more complex and clunky!
- Nonetheless, IAMs are an essential part of the toolkit used by IPCC and similar other orgs to **inform policymakers** on the effect of policies on society and climate change.
- They are wildly computationally expensive, which results in them being **difficult to probe**. They also often **behave in unexpected ways**.
- Here, we probe a relatively simple IAM (called AYS) with **Reinforcement Learning** (RL) agents to analyse its behavior. We also use these agents to **explore potential solutions** to the model.

AYS: the environment

AYS model is a 3-dimensional dynamical model governed by three variables:

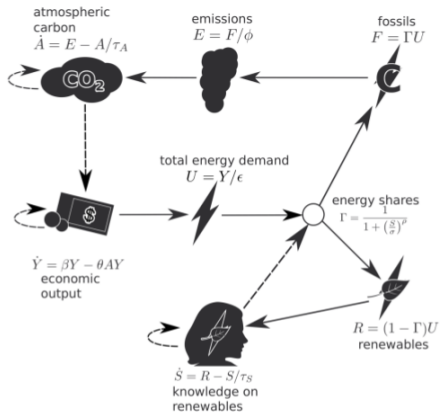
A: Atmospheric Carbon

Y: Economic output

S: Renewable energy knowledge

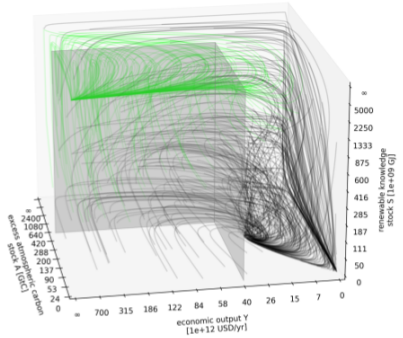
All the variables depend on each other with differential equations.

The parameters of the model are **fitted to real world values**.



The AYS model from Kittel et al.

Action	Effect	Interpretation
No-operation (noop)	System evolves as is	Inaction
Energy Transition (ET)	Renewable energy becomes cheaper	Subsidies and/or R&D investment
Degrowth (DG)	Economy growth slows down	High taxes and fines
Combination (ET+DG)	Combination of both actions above	Strong policymaker intervention



AYS (phase space)

Lines == trajectories

Hyperplanes are *planetary boundaries*

Endpoints/attractors

Green end: **low carbon / high growth** scenario - the agents' goal.

Black end: a **high carbon / low growth** scenario.

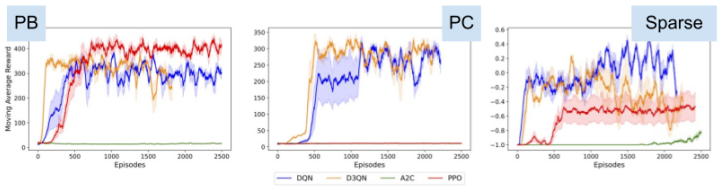
Reward functions

Planetary Boundary (PB): the distance between the current state and the Planetary Boundaries. *Maximising distance to undesirable states*.

Policy Cost (PC): adds an action-dependent cost to the PB reward, simulating the real-world *cost of implementing & maintaining any significant shift in policy*.

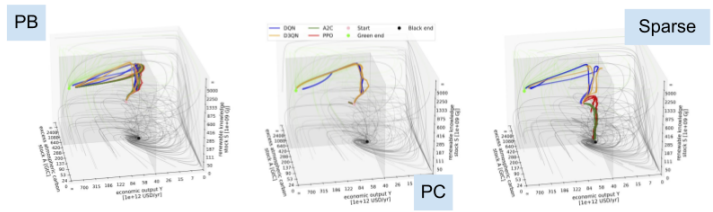
Sparse: Gives a *reward of 1 for reaching the goal or -1 for failing*. During the episode all other actions yield a reward of 0.

Results

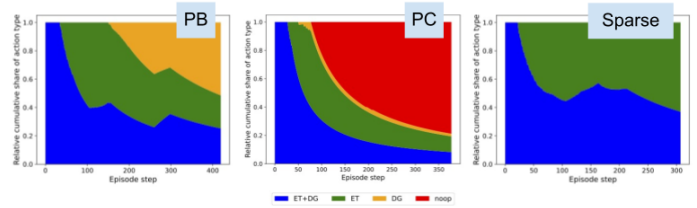


- The 4 learning algorithms have **different learning patterns** for **different reward functions**.
- Off-policy agents perform more consistently across all reward functions.
- Sparse signal impedes the ability for the agents to learn effectively.
- Needs for correct incentive structures for the agents to learn successful policies!

There are many different possible pathways for the agents to succeed!

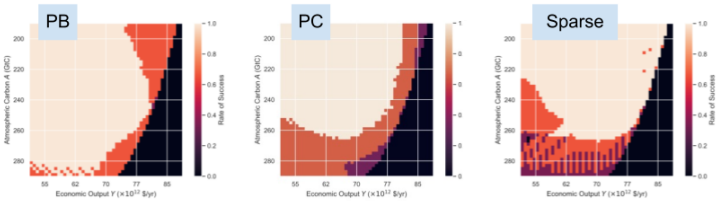


Sample trajectories of agents trained with different reward functions.



Relative cumulative share of actions taken over time by trained DQN agents trained with 3 different reward functions

Under any reward function, the combined action is the preferred for the first 23 steps of the episode is ET+DG.



Average rate of reaching the green point of 3 seeds of the trained DQN agents trained with 3 different reward functions given different initialisation states.

Any initial state with high economic output and high carbon is **impossible to solve for the agents**.

Conclusion

- Agents learn **different optimal pathways** that reach the goal.
- Depending on the reward function some agents perform better than others under the same episodic training budget.
- A change in reward function will impact the resultant agent policy - as well as its performance - significantly: this shows the importance of constructing the **correct incentive structure for climate change policy.**
- All the successful pathways rely on the agent using the **highest impact action early on** in the episode. This draws a parallel to the concept of **early action**.
- Reinforcement Learning can be used as a tool for **analyzing and debugging IAMs.**

References

Strnad, F.M. et al. (2019) 'Deep reinforcement learning in world-earth system models to discover Sustainable Management Strategies', Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(12).

Nitzbon, J., Heitzig, J. and Parlitz, U. (2017) 'Sustainability, collapse and oscillations in a simple world-earth model', Environmental Research Letters, 12(7), p. 074020.

Moore, F.C., Lacasse, K., Mach, K.J. *et al.* Determinants of emissions pathways in the coupled climate–social system. *Nature* 603, 103–111 (2022).

[1]Carbon Re, [2]University College London