Climate Surrogates for Scalable Multi-Agent Reinforcement Learning: A Case Study with CICERO-SCM

Extended Abstract

Oskar Bohn Lassen
Technical University of Denmark
Kongens Lyngby, Denmark
obola@dtu.dk

Serio Angelo Maria Agriesti
Technical University of Denmark
Kongens Lyngby, Denmark
samaa@dtu.dk

Filipe Rodrigues
Technical University of Denmark
Kongens Lyngby, Denmark
rodr@dtu.dk

Francisco Camara Pereira
Technical University of Denmark
Kongens Lyngby, Denmark
camara@dtu.dk

ABSTRACT

We achieve a $> 100\times$ speedup in multi-agent climate policy reinforcement learning (RL) by replacing the CICERO-SCM climate simulator with a high-fidelity surrogate model. This surrogate captures multi-gas climate dynamics with near-simulator accuracy (global-mean temperature RMSE $\approx$ 0.0004K) while running approximately 1000× faster per simulation step. Bypassing the core computational bottleneck, the surrogate enables regional agents to learn climate policies under multi-gas dynamics in scenarios where the original simulator is intractable. Our approach preserves policy fidelity (converging to the same solutions as the original simulator) and unlocks large-scale multi-agent experiments across alternative climate-policy regimes with high-fidelity climate response[1].

KEYWORDS

Surrogate modeling; Climate simulation; Multi-agent reinforcement learning; Climate policy analysis

ACM Reference Format:
Oskar Bohn Lassen, Serio Angelo Maria Agriesti, Filipe Rodrigues, and Francisco Camara Pereira. 2026. Climate Surrogates for Scalable Multi-Agent Reinforcement Learning: A Case Study with CICERO-SCM: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/RJBJ9974

1 INTRODUCTION

High-fidelity climate models provide detailed projections but are notoriously slow [1]. Simpler climate models (SCMs) like MAGICC, FaIR or CICERO-SCM run much faster, making them popular for policy analysis [3, 5, 11, 13, 20]. Integrated assessment models (IAMs) are widely used to evaluate climate policies but typically rely on equilibrium assumptions and aggregate decision-making, which limit agent heterogeneity and dynamic interactions [4, 8, 14, 18]. Multi-agent reinforcement learning (MARL) has been proposed for climate-economy analysis to accommodate heterogeneous agents, non-linear dynamics, and remove bound rational decision-making - but even SCMs can become a bottleneck when embedded in MARL environments that require millions of simulator calls [15]. Hence, prior studies have used highly simplified climate dynamics (e.g. a single CO_2 curve), limiting realism and policy exploration [16, 21, 22]. To bridge this gap, we integrate a learned climate surrogate into the environment, aiming to retain the multi-gas fidelity of a complex SCM while drastically speeding up training. Our framework achieves $> 100\times$ faster training without sacrificing optimal policy outcomes, enabling large-scale MARL experiments with high-fidelity climate responses.

2 APPROACH

We use CICERO-SCM as the climate engine, f_{SCM}, in our MARL experiment [17]. CICERO-SCM is a reduced-complexity model that maps annual emissions of $|G| = 40$ greenhouse gases to global mean temperature change $\Delta T(t)$. It requires ~0.4 seconds per call, which is too slow for MARL training. Hence, we develop a fast surrogate model f_θ to approximate f_{SCM}. We generated an ensemble of 20,000 emission pathways by perturbing the SSP2-4.5 baseline scenario (2015–2075), varying year-over-year growth rates for key gases (CO_2 (fossil and land-use), CH_4, N_2O, SO_2) within ±7.5% aligning with primary interventions according to IPCC Sixth Assessment Report (AR6) [2, 6, 12]. These scenarios are run through CICERO-SCM to provide training data for a recurrent neural network (RNN) surrogate. The RNN takes a window of recent emissions as input and outputs $\Delta T(t)$, effectively emulating the simulator's one-year update.

We implemented a multi-agent climate-economic game as a Markov game with N regional agents, each learning policies over a finite horizon H by interacting with a shared climate and impact environment. Each agent i selects actions $a_{i,t}$ from a discrete policy space and observes centralized state variables o_t. The actions are effort levels of energy decarbonization, methane abatement, agricultural and land-use measures, and preventive investment (representing climate adaptation measures). The joint actions determine

---

[1] An extended version of this work is available in [9] and code can be accessed here.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. https://doi.org/10.65109/RJBJ9974

Table 1: Surrogate performance on held-out test data and policy-induced trajectories and acceleration of inference speed and MARL environment step in scenario (i). Speed-up is measured relative to CICERO-SCM.

	Test data performance & inference speed			**Policy-induced performance & MARL speed**			
Climate engine	**Test data (RMSE, R^2)**	**Mean inference [s] (CPU/GPU)**	**Speed-up (CPU/GPU)**	**Scenario (i) (RMSE, rank-τ)**	**Scenario (ii) (RMSE, rank-τ)**	**Mean env-step [s]**	**Speed-up**
CICERO–SCM	–	464.4×10^{-3} / –	–	–	–	217.7×10^{-3}	–
LSTM	$(4.7 \times 10^{-4}, 0.99)$	1.1×10^{-3} / 0.4×10^{-3}	442× / 1161×	$(5.9 \times 10^{-4}, 0.996)$	$(3.2 \times 10^{-4}, 0.990)$	1.6×10^{-3}	137×
GRU	$(3.7 \times 10^{-4}, 0.99)$	2.3×10^{-3} / 0.4×10^{-3}	202× / 1161×	$(3.9 \times 10^{-4}, 0.996)$	$(2.0 \times 10^{-4}, 0.997)$	1.6×10^{-3}	137×
TCN	$(6.8 \times 10^{-4}, 0.99)$	3.3×10^{-3} / 1.3×10^{-3}	140× / 357×	$(21.1 \times 10^{-4}, 0.994)$	$(10.3 \times 10^{-4}, 0.982)$	4.5×10^{-3}	49×

multi-gas emissions E_t, which drive temperature change ΔT_t via the climate model (surrogate f_θ or simulator f_{SCM}). Agents receive rewards based on mitigation costs, adaptation investments, and climate damages, and aim to maximize $J_i(\theta_i) = \mathbb{E}_{\tau\sim\pi_{\theta_i}}\left[\sum_{t=0}^{H-1} \gamma^t\, r_i(t)\right]$.

Scenario (i) features $N = 4$ homogeneous agents with identical characteristics and a single mitigation lever (energy decarbonization), enabling training with both the surrogate and simulator. Scenario (ii) involves $N = 10$ heterogeneous agents with varied emission shares, cost sensitivities, and multiple levers (decarbonization, methane abatement, land-use, adaptation), yielding a high-dimensional policy space. Training with the simulator becomes intractable in this setting, so agents learn with the surrogate.

Policy consistency criterion. An ideal surrogate should induce the same optimal policies as the climate simulator. Principles on policy consistency in model-based RL are defined as:

$$\text{sign}\left[\Delta J_{f_{\text{NET}}}(\pi_1,\pi_2)\right] \approx \text{sign}\left[\Delta J_{f_{\text{SCM}}}(\pi_1,\pi_2)\right], \quad \forall\, \pi_1,\pi_2 \in \Pi \tag{1}$$

$$\nabla_\theta J_{f_{\text{NET}}}(\pi_\theta) \approx \nabla_\theta J_{f_{\text{SCM}}}(\pi_\theta), \qquad \forall\, \theta \in \mathcal{N}(\theta^\star_{\text{SCM}}) \tag{2}$$

where $J_f(\pi)$ is the expected return (cumulative discounted reward) under climate model f and policy π, and θ^*_{SCM} are the parameters of the optimal policy for the simulator [7, 10, 19]. The first condition requires f_θ and f_{SCM} to induce the same preference ordering over policies, and the second ensures that the local gradient of the return is aligned around the optimum - together implying convergence to the same equilibrium policy. Directly verifying these conditions would require training both models, which is often intractable. We therefore propose an empirical consistency check: after training with the surrogate, we replay a set of N policy-induced emission trajectories through the original simulator. We then compare the surrogate and simulator outcomes on these trajectories via two metrics: (1) RMSE between the temperature responses $\Delta T_{f_\theta}(t)$ and $\Delta T_{f_{\text{SCM}}}(t)$; and (2) Kendall's τ rank-correlation between the returns (cumulative temperature-impact rewards) under the surrogate and simulator. A high $\tau \approx 1$ indicates that the surrogate preserves the ranking of policies by performance, satisfying the consistency criterion in equation (1). This method provides a tractable check of policy consistency under the following assumption:

$$\left\|\mathcal{S} - \bar{\mathcal{S}}\right\| \;\to\; 0 \quad \text{as} \quad \varepsilon \to 0 \tag{3}$$

where ϵ is a metric of the error between $\Delta T_{f_\theta}(t)$ and $\Delta T_{f_{\text{SCM}}}$ and $\mathcal{S}$ and $\bar{\mathcal{S}}$ are the sets of emission trajectories visited during policy optimization with the simulator and surrogate respectively.

3 RESULTS & DISCUSSION

Table 1 (left columns) quantifies the surrogate's performance with a test-set RMSE $\approx 4 \times 10^{-4}$ K. Each one-year climate step can be computed in ~0.0004 seconds on a GPU (vs. 0.4 s for CICERO-SCM), yielding about 1000× speed-up in per-step climate inference. This translates into over 100× faster end-to-end training when the surrogate is used in the MARL loop. The policy consistency results (right columns of Table 1) strongly support our approach. In the 4-agent tractable scenario (i), we could directly compare the learned policies and see that training with f_θ converges to the same policies as training with f_{SCM}. This was also supported by the Kendall's $\tau \approx 0.99$ between the surrogate-induced returns and simulator returns which indicates almost identical ordering of policies. In the 10-agent scenario (ii), direct simulator training was infeasible, but our replay evaluation reveals that the surrogate's learned policy trajectories still yield extremely low error (even lower RMSE than in scenario i) and a high rank-correlation ($\tau \approx 0.99$) with the simulator. This implies that the surrogate-induced policies would be consistent with the ones learned using the simulator. In other words, the surrogate preserves the policy fidelity: it incurs only negligible return discrepancies along relevant trajectories, thus maintaining the optimal policy. Overall, the surrogate-integrated MARL training achieves massive speed-ups without compromising the climate-policy outcomes.

4 CONCLUSION

We trained a high-fidelity climate surrogate and embedded it into a climate-economic MARL environment achieving near-perfect accuracy and a speed-up of > 1000× for one-step inference and > 100× for total MARL training. We proposed an empirical policy-consistency check using simulator replay and rank-correlation which indicated that using the surrogate did not alter the agents' learned policies. Together, these results demonstrate that high-fidelity, multi-gas climate models can be faithfully approximated and deployed in RL environments, removing a major computational barrier to scalable climate-policy analysis. This opens the door to MARL studies that explore richer climate dynamics, multiple gases, and many agents - scenarios previously intractable with direct simulation.

ACKNOWLEDGMENTS

The work presented in this article is supported by Novo Nordisk Foundation grant NNF23OC0085356.

REFERENCES

[1] Gordon B. Bonan and Scott C. Doney. 2018. Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models. *Science* 359, 6375 (2018), eaam8328. https://doi.org/10.1126/science.aam8328

[2] Leon Clarke, Y.-M. Wei, Ana De La Vega Navarro, Amit Garg, and et al. 2022. Energy Systems. In *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA. https://doi.org/10.1017/9781009157926.008

[3] Kalyn Dorheim, Robert Link, Corinne Hartin, Ben Kravitz, and et al. 2020. Calibrating Simple Climate Models to Individual Earth System Models: Lessons Learned From Calibrating Hector. *Earth and Space Science* 7, 11 (2020), e2019EA000980. https://doi.org/10.1029/2019EA000980 e2019EA000980 10.1029/2019EA000980.

[4] Oliver Fricko, Petr Havlik, Joeri Rogelj, Zbigniew Klimont, and et al. 2017. The marker quantification of the Shared Socioeconomic Pathway 2: A middle-of-the-road scenario for the 21st century. *Global Environmental Change* 42 (2017), 251–267. https://doi.org/10.1016/j.gloenvcha.2016.06.004

[5] Jan S. Fuglestvedt and Terje K. Berntsen. 1999. *A Simple Model for Scenario Studies of Changes in Global Climate: Version 1.0.* CICERO Working Paper 1999:02. Center for International Climate and Environmental Research (CICERO).

[6] Intergovernmental Panel on Climate Change (IPCC). 2022. Summary for Policymakers. In *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3–33. https://doi.org/10.1017/9781009325844.001

[7] Michael Janner, Justin Fu, Marvin Zhang, and Sergey Levine. 2021. When to Trust Your Model: Model-Based Policy Optimization. arXiv:1906.08253 [cs.LG]

[8] Konstantinos Koasidis, Alexandros Nikas, and Haris Doukas. 2023. Why integrated assessment models alone are insufficient to navigate us through the polycrisis. *One Earth* 6, 3 (2023), 205–209. https://doi.org/10.1016/j.oneear.2023.02.009

[9] Oskar Bohn Lassen, Serio Angelo Maria Agriesti, Filipe Rodrigues, and Francisco Camara Pereira. 2025. Climate Surrogates for Scalable Multi-Agent Reinforcement Learning: A Case Study with CICERO-SCM. arXiv:2510.07971 [cs.LG] https://arxiv.org/abs/2510.07971

[10] Yecheng Jason Ma, Kausik Sivakumar, Jason Yan, Osbert Bastani, and et al. 2023. Learning Policy-Aware Models for Model-Based Reinforcement Learning via Transition Occupancy Matching. In *Proceedings of The 5th Annual Learning for Dynamics and Control Conference (Proceedings of Machine Learning Research, Vol. 211)*. PMLR, 259–271.

[11] M. Meinshausen, S. C. B. Raper, and T. M. L. Wigley. 2011. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration. *Atmospheric Chemistry and Physics* 11, 4 (2011), 1417–1456. https://doi.org/10.5194/acp-11-1417-2011

[12] G.-J. Nabuurs, R. Mrabet, A. Abu Hatab, M. Bustamante, and et al. 2022. Agriculture, Forestry and Other Land Uses (AFOLU). In *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK and New York, NY, USA. https://doi.org/10.1017/9781009157926.009

[13] Zebedee R. J. Nicholls, M. Meinshausen, J. Lewis, R. Gieseke, and et al. 2020. Reduced Complexity Model Intercomparison Project Phase 1: introduction and evaluation of global-mean temperature response. *Geoscientific Model Development* 13, 11 (2020), 5175–5190. https://doi.org/10.5194/gmd-13-5175-2020

[14] William Nordhaus. 2018. Projections and Uncertainties about Climate Change in an Era of Minimal Climate Policies. *American Economic Journal: Economic Policy* 10, 3 (August 2018), 333–60. https://doi.org/10.1257/pol.20170046

[15] Georgios Papoudakis, Filippos Christianos, Lukas Schäfer, and Stefano V. Albrecht. 2021. Benchmarking Multi-Agent Deep Reinforcement Learning Algorithms in Cooperative Tasks. arXiv:2006.07869 [cs.LG]

[16] James Rudd-Jones, Mirco Musolesi, and María Pérez-Ortiz. 2025. Multi-Agent Reinforcement Learning Simulation for Environmental Policy Synthesis. arXiv:2504.12777 [cs.MA]

[17] Maria Sandstad, B. Aamaas, A. N. Johansen, M. T. Lund, and et al. 2024. CICERO Simple Climate Model (CICERO-SCM v1.1.1) – an improved simple climate model with a parameter calibration tool. *Geoscientific Model Development* 17, 17 (2024), 6589–6625. https://doi.org/10.5194/gmd-17-6589-2024

[18] Ivan Savin, Felix Creutzig, Tatiana Filatova, Joël Foramitti, and et al. 2023. Agent-based modeling to integrate elements from different disciplines for ambitious climate policy. *WIREs Climate Change* 14, 2 (2023), e811. https://doi.org/10.1002/wcc.811

[19] Jian Shen, Hang Lai, Minghuan Liu, Han Zhao, and et al. 2023. Adaptation Augmented Model-based Policy Optimization. *Journal of Machine Learning Research* 24, 218 (2023), 1–35.

[20] Christopher J. Smith, P. M. Forster, M. Allen, N. Leach, and et al. 2018. FAIR v1.3: a simple emissions-based impulse response and carbon cycle model. *Geoscientific Model Development* 11, 6 (2018), 2273–2297. https://doi.org/10.5194/gmd-11-2273-2018

[21] Felix M. Strnad, Wolfram Barfuss, Jonathan F. Donges, and Jobst Heitzig. 2019. Deep reinforcement learning in World-Earth system models to discover sustainable management strategies. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29, 12 (12 2019), 123122. https://doi.org/10.1063/1.5124673

[22] Tianyu Zhang, Andrew Williams, Soham Phade, Sunil Srinivasa, and et al. 2022. AI for Global Climate Cooperation: Modeling Global Climate Negotiations, Agreements, and Long-Term Cooperation in RICE-N. arXiv:2208.07004 [cs.LG]