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

Oskar Bohn Lassen[1] , **Serio Agriesti**[1] , **Filipe Rodrigues**[1] , **Blaz Kurnik**[2] , **Francisco Pereira**[1]
[1]Technical University of Denmark
[2]European Environmental Agency
{obola, samaa, rodr}@dtu.dk, blaz.kurnik@eea.europa.eu, camara@dtu.dk

Abstract

Climate policy analysis requires models that capture multi-gas climate effects, but such models are too slow to embed in reinforcement learning loops at scale. In collaboration with a pan-European public-sector environmental agency, we develop a multi-agent reinforcement learning (MARL) framework that integrates a higher-fidelity climate surrogate as the environment transition, enabling regional agents to learn policies under multi-gas dynamics. We train a recurrent surrogate on $20{,}000$ multi-gas emission pathways to emulate CICERO-SCM. The surrogate achieves near-simulator accuracy (global-mean temperature RMSE $\approx 4\times10^{-4}$ K) with $\sim 1000\times$ faster one-step inference and yields $> 100\times$ end-to-end MARL training speed-up. We show policy agreement with the simulator in tractable settings and propose a replay- and rank-consistency test (Kendall's τ) for assessing policy fidelity when simulator-in-the-loop training is infeasible. This enables large-scale multi-agent policy experiments while retaining high-fidelity multi-gas climate response.

1 Introduction

Climate modeling provides the backbone for climate policy exploration, but there is an inherent trade-off between model fidelity and computational cost. State-of-the-art Earth System Models (ESMs) simulate climate processes in great detail [Bonan and Doney, 2018], but a single ESM run can take days to weeks on high-performance computers. This extreme cost severely limits how many scenarios or policy strategies can be evaluated, motivating the use of simpler models for many applications [Romero-Prieto *et al.*, 2025].

The climate science community relies on reduced-complexity Simple Climate Models (SCMs) to emulate climate responses at far lower computational cost. SCMs are compact models (often energy-balance models with simplified ocean and carbon-cycle components) calibrated to reproduce the behavior of more complex ESMs [Dorheim *et al.*, 2020]. For example, widely used SCMs like MAGICC [Meinshausen *et al.*, 2011] and CICERO-SCM [Fuglestvedt and Berntsen, 1999] provide tractable climate projections using simplified physics, while the FaIR model takes an impulse-response approach [Smith *et al.*, 2018; Nicholls *et al.*, 2020]. These models sacrifice some process-level detail for high computational efficiency. Because SCMs run orders of magnitude faster than ESMs, they are widely used in applications that require large ensembles or iterative evaluations, such as probabilistic climate projections or integrated assessment models (IAMs) [Romero-Prieto *et al.*, 2025].

IAMs couple socio-economic systems with a climate module to assess mitigation and impact pathways, translating climate outcomes into economic metrics. Pioneering IAMs like DICE and RICE used highly simplified climate modules in an optimal growth framework [Nordhaus, 1992; Nordhaus, 1993; Nordhaus, 2014; Nordhaus and Yang, 1996; Nordhaus, 2010; Nordhaus, 2018]. Building on this foundation, more detailed IAM frameworks (e.g., REMIND-MAgPIE, MESSAGE-GLOBIOM, WITCH, IMAGE) introduced additional realism (e.g. technological detail, land-use, energy systems) [Kriegler *et al.*, 2017; Fricko *et al.*, 2017a; Emmerling *et al.*, 2016; Stehfest *et al.*, 2014], often exploring socio-economic scenario families (SSPs) [Riahi *et al.*, 2017]. In all cases, IAMs rely on fast climate simulators like MAGICC [Meinshausen *et al.*, 2011] because ESMs are too slow for large scenario ensembles or iterative policy optimization loops. However, traditional IAMs typically model the world as a few aggregate regions optimizing an equilibrium trajectory under strong foresight. This aggregation and reliance on optimal-control or game-theoretic assumptions can mask heterogeneity in preferences, adaptive behaviors, and path-dependent collective dynamics. These limitations have prompted interest in more flexible, simulation-based approaches [Koasidis *et al.*, 2023; Savin *et al.*, 2023].

Reinforcement learning (RL), especially multi-agent RL (MARL), has been proposed as an alternative framework for studying climate-economy interactions [Strnad *et al.*, 2019; Rudd-Jones *et al.*, 2025]. In an MARL formulation, multiple agents (e.g. countries) make simultaneous decisions and learn strategies through repeated interaction in a simulated environment, rather than assuming a globally optimized equilibrium path. This approach accommodates heterogeneous agents, non-linear dynamics, and bounded rationality. An early attempt to apply MARL in an IAM context is RICE-N [Zhang *et al.*, 2022], which replaced RICE's decision-making with learning agents. However, RICE-N used an overly simplified climate module and limited cooperation mechanisms, restricting its realism and the policies that could emerge. Subsequent works began to address these gaps:[Rudd-Jones *et al.*, 2024] incorporated more sophisticated cooperation and coalition formation, and [Heitzig *et al.*, 2023] examined the impact of commitment strategies on climate outcomes. The recent JUSTICE framework [Biswas *et al.*, 2025] went further by integrating the FaIR climate model into a multi-objective MARL game, improving the fidelity of climate dynamics in the learning environment. Nonetheless, even JUSTICE only allowed control of CO_2 with a low-dimensional action space and a small number of agents, limiting exploration of complex policies and undercutting the benefits of a detailed climate response.

In reality, different greenhouse gases and aerosols have widely varying lifetimes and warming effects. Long-lived gases like CO_2 and N_2O persist for centuries (driving long-term warming), methane lasts around a decade (strong near-term warming), and short-lived pollutants (e.g. aerosols) persist for months and often have a cooling effect. Mitigation actions also influence these species heterogeneously: for example, decarbonizing the energy sector reduces CO_2 emissions but also lowers SO_2 aerosol emissions (removing some cooling and causing a short-term warming), while halting deforestation cuts CO_2 (land-use) and CH_4 without a similar short-term penalty. Thus, MARL climate games need a climate engine that responds to multi-gas emission patterns rather than just aggregate CO_2. To date, however, MARL studies have relied on oversimplified climate dynamics and severely restricted action spaces, mainly due to computational constraints.

Scalability is a critical issue: MARL experiments often require 10^7 environment interactions for agents to learn effective policies [Papoudakis *et al.*, 2021]. Even an SCM taking $\sim\ 0.4$ seconds per step can become a bottleneck when millions of calls are needed. This is a key reason why prior MARL studies have not used complex or multi-gas climate models - doing so would make training intractable. One promising approach to embed higher-fidelity climate dynamics into MARL frameworks without incurring prohibitive cost is the use surrogate models. Machine-learning based surrogates have accelerated expensive components of ESMs by orders of magnitude [Lu and Ricciuto, 2019], and deep neural networks can mimic short-term climate projections with high speed and reasonable accuracy [Weber *et al.*, 2020]. However, most prior climate surrogates target offline surrogate modeling, whereas in this setting, a surrogate must become a component of a MARL environment transition function, where small dynamics errors can alter exploration, coordination, and even the learned equilibrium.

In this paper, we extend the realism of MARL climate-policy games by embedding a more complex, multi-gas climate model into the environment while keeping training tractable. First, we propose a modular framework for substituting a surrogate model in place of the climate module within a climate-economy MARL environment. Second, we introduce a recurrent neural network surrogate for CICERO-SCM and describe our MARL experimental setup, including a method to evaluate policy consistency when running the simulator is infeasible. Finally, we present results demonstrating that our surrogate dramatically accelerates MARL training ($> 100\times$ speed-up) without sacrificing policy fidelity.

The work has been developed in collaboration with the European Environmental Agency where a domain expert have engaged with us to align the problem setting with policy-relevant use cases and operational constraints (e.g., realism of emission pathways and controllable levers) and have provided feedback on the experimental protocol and interpretation of results.

2 Conceptual Framework

According to the IPCC Sixth Assessment Report [Intergovernmental Panel on Climate Change (IPCC), 2022; Nabuurs *et al.*, 2022; Clarke *et al.*, 2022], key intervention areas in climate policy include: (i) decarbonizing the energy sector (e.g. coal phase-out, renewable expansion), (ii) targeting methane abatement (e.g. waste management, leak mitigation), (iii) improving agricultural and land-use practices (e.g. reduced deforestation, fertilizer efficiency), and (iv) investing in adaptation or preventive measures that reduce realized damages. These policy levers affect multiple greenhouse gases, each with different climate effects. Ideally, the resulting emission changes should feed into a high-fidelity climate model to provide a realistic estimate of temperature increase, which then propagates into region-specific damage estimates.

However, there is a trade-off between environment complexity and tractability: simplifying the climate model makes experiments feasible but limits the policies and dynamics that can be explored. We propose a modular framework in which complex climate components can be replaced by fast surrogates. We divide the environment conceptually into three modules:

1. **Emissions Module:** maps agent actions to multi-gas emissions
2. **Climate Module:** maps emissions to climate response, e.g., temperature change
3. **Impact Module:** maps climate outcomes and actions to costs or damages

As illustrated in Figure 1, one can seamlessly substitute a higher-fidelity climate model f_{SCM} with a learned surrogate model f_θ, without altering any other part of the environment. The surrogate emulates the original model's behavior, preserving the multi-gas climate response fidelity while enabling the use of a high-speed, hardware-optimized model in the RL training loop. This modular design also provides flexibility to increase fidelity in other parts of the environment without changing the learning algorithm. For instance, the impact module could be a simple global damage function or a detailed model of local impacts (e.g., regional sea-level rise or crop yield changes) without requiring any modifications to the agents' decision-making loop.

In the remainder of this paper, we instantiate this framework using CICERO-SCM as our high-fidelity climate engine and a recurrent neural network (RNN) as the surrogate emulator. We then integrate the surrogate into a multi-agent

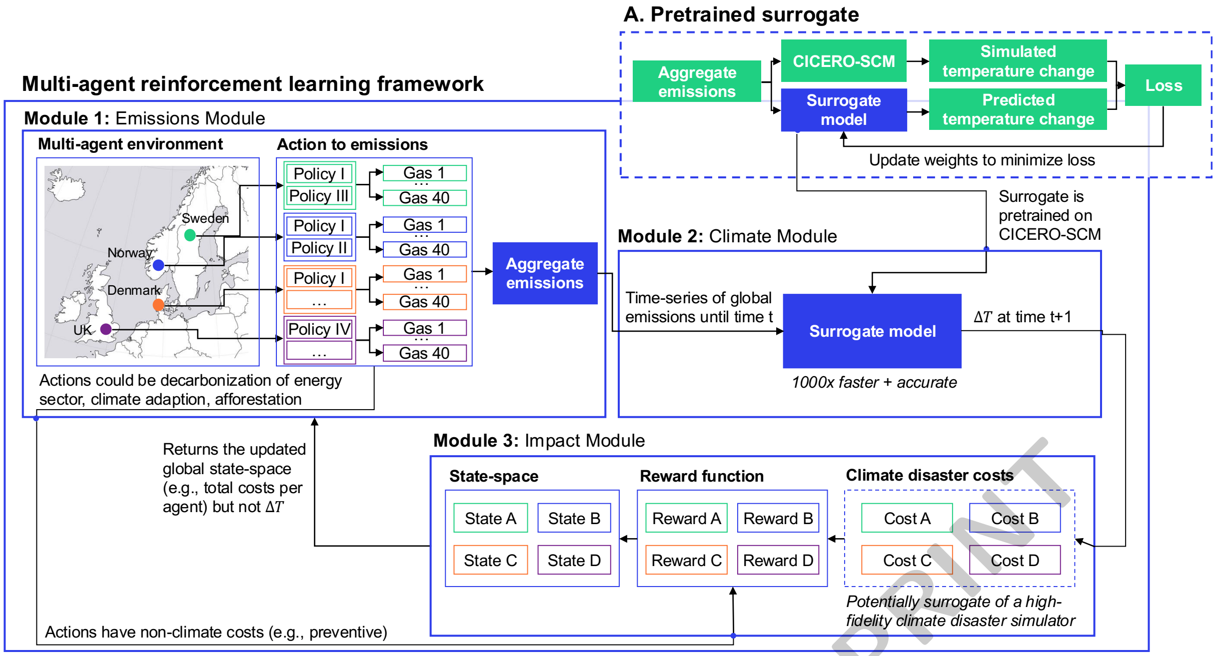

Figure 1: Framework for integrating climate surrogates into MARL. In Module 1, agents choose policies that result in emissions, which in Module 2 are mapped into temperature change by a pretrained surrogate and in Module 3 converted into costs.

climate-economic environment and demonstrate the resulting speed-ups and policy consistency.

3 Methodology

In this section, the methodological steps adopted to design a surrogate for CICERO-SCM and embed it into the MARL framework are described.

3.1 Climate Dynamics Engine: CICERO-SCM

CICERO-SCM (v1.1.1) [Sandstad *et al.*, 2024] is a reduced-complexity global climate model that maps multi-gas emission trajectories to global-mean surface air temperature. Let G be the set of gases $((|\mathcal{G}| = 40))$ used in the model, and define the global emissions vector $E(t) = [E_g(t)]_{g\in\mathcal{G}} \in \mathbb{R}^{|\mathcal{G}|}$ and the emissions history $E_{1:t} = [E(\tau)]_{\tau=1}^{t} \in \mathbb{R}^{t\times|\mathcal{G}|}$.

At each year t, the simulator produces a temperature change $\Delta T(t) = f_{\text{SCM}}(E_{1:t})$, where $\Delta T(t)$ is the global mean surface air temperature change in year t. Internally, CICERO-SCM couples a carbon-cycle module for CO_2, exponential-decay models for other long-lived gases (e.g., CH_4, N_2O,), and an upwelling-diffusion energy-balance model that translates total radiative forcing into temperature change [Etminan *et al.*, 2016]. Its parameters are calibrated to ESM outputs and observations, yielding a physically consistent climate response (Appendix A.2). However, each call still takes ~ 0.4 s, which would make millions of simulator steps (as required in MARL) intractable.

3.2 Surrogate Model of CICERO-SCM

Generation of Emission Trajectories for Model Training
For surrogate training, we generated an ensemble of $S =$ 20,000 policy-relevant emission trajectories by randomly perturbing the year-over-year growth rates of a baseline scenario (SSP2-4.5 [Fricko *et al.*, 2017b]) over the period 2015-2075. We apply smoothing to these perturbations to produce realistic trajectories that mimic the smoother patterns of learned policies. Specifically, we first compute the baseline annual growth factor for each gas g:

$$\delta_g^{\text{base}}(t) = \frac{E_g^{\text{base}}(t)}{E_g^{\text{base}}(t-1)} \tag{1}$$

For each scenario s, we draw multiplicative factors $\tilde{\zeta}_g^s(t) \sim \mathcal{U}(\ell_g, u_g)$ and then smooth them via an exponential moving average:

$$\zeta_g^s(t) = \left(\zeta_g^s(t-1)\right)^{\alpha} \left(\tilde{\zeta}_g^s(t)\right)^{1-\alpha}, \quad \zeta_g^s(2015) = 1 \tag{2}$$

These factors perturb the baseline growth, $\delta_g^s(t) = \delta_g^{\text{base}}(t)\,\zeta_g^s(t)$. This defines scenario-specific emissions recursively: for $t \le 2015$, $E_g^s(t) = E_g^{\text{base}}(t)$ (using historical emissions); for $t \ge 2016$, $E_g^s(t) = E_g^s(t-1)\,\delta_g^s(t)$.

We focus on a subset $\mathcal{C} \subset \mathcal{G}$ of five controllable gas species to be controlled in the MARL experiment (Section 3.3): CO_2 from fossil fuels (FF), CO_2 from land-use (AFOLU), CH_4, N_2O, and SO_2. Perturbations are applied only to these gases by setting $(\ell_g, u_g) = (0.925, 1.075)$ for $g \in C$ (allowing $\pm 7.5\%$ annual growth variation) and $(\ell_g, u_g) = (1, 1)$ for $g \notin C$ (those emissions follow the baseline). We choose $\pm 7.5\%$ to intentionally exceed the MARL agents' maximum per-year control range ($\pm 5\%$; Appendix D, Tables A.3-A.4), so that all policy-induced emission trajectories in the intended MARL setting remain within the surrogate's training support.

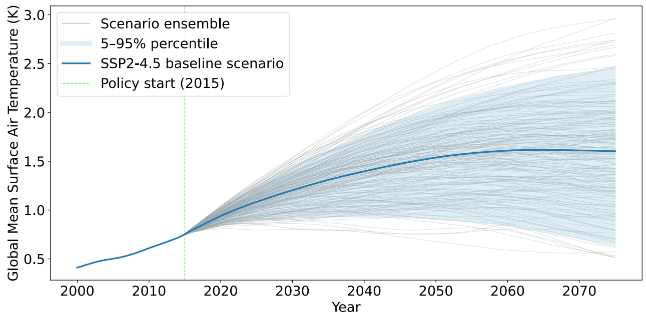

Figure 2: Global mean surface air temperature change for the generated emission trajectories.

Thus, the surrogate is designed for accuracy on this feasible policy-induced emission space, rather than for unconstrained extrapolation far outside these bounds. These perturbations are sampled independently across gases to ensure coverage of the controllable action subspace which avoids under-representing low-magnitude gases that could otherwise be overshadowed by correlated variations. Running CICERO-SCM on all 20,000 scenarios yields an ensemble of temperature trajectories (Figure 2). The spread of global temperature outcomes is narrow in the near term (since all scenarios share the same historical emissions) but grows substantially by 2075 as the scenarios diverge. The unperturbed SSP2-4.5 baseline lies near the median of the ensemble, indicating that our scenario generation spans futures consistent with that reference while covering substantial variability in warming outcomes.

For each scenario s and year t, let $E^s(t) = [E^s_g(t)]_{g\in\mathcal{G}}$ be the full emission vector and $E^s_{\mathcal{C}}(t) = [E^s_g(t)]_{g\in\mathcal{C}}$ the vector of emissions for the controllable gases. We use the paired data $\{E^s_{\mathcal{C}}(t), \Delta T^s(t)\}$ to train the surrogate:

$$X^s(t) = \left[E^s_{\mathcal{C}}(t-W),\ \ldots,\ E^s_{\mathcal{C}}(t)\right], \quad y^s(t) = \Delta T^s(t) \quad (3)$$

where W is the length of the temporal window. We choose $W = 65$ years so that the input covers sufficient historical context, to capture slow climate responses and long-lived gas effects, while remaining computationally manageable. We generate one sample for each scenario and each target year $t \in \{2015, \ldots, 2075\}$ (excluding years before 2015 since all scenarios share identical emissions). This yields 61 samples per scenario, for a total of about 1.22 million training examples. The data are split by scenario into 70% training, 15% validation, and 15% testing, ensuring that no scenario's future data appear in its training set.

Architecture of RNN-based Surrogate Model We build a recurrent neural network surrogate comprising three modules: (i) an RNN-based encoder, (ii) a skip-connection that feeds current-year emissions directly to the predictor, and (iii) a two-layer feedforward prediction head. The RNN encoder processes the historical emissions window $[x_{t-W}, \ldots, x_{t-1}]$, where $x_\tau \in \mathbb{R}^{|\mathcal{C}|}$ is the controllable-gas emissions at time τ, and produces a hidden state h_t summarizing the past dynamics. We experimented with different sequence models for this encoder, including Long Short-Term Memory (LSTM) [Hochreiter and Schmidhuber, 1997], Gated Recurrent Unit (GRU) [Chung *et al.*, 2014], and a Temporal Convolutional Network (TCN) [Lea *et al.*, 2016] (which replaces recurrence with convolutions while preserving the input-output structure). These architectures are well suited to the annual, low-dimensional emission sequences considered here, while keeping inference fast inside the MARL loop. A skip-connection then concatenates the current-year emissions x_t with the encoded history to form $z_t = [h_t; x_t]$, ensuring that recent emission changes are directly available to the predictor. Finally, the prediction head - a two-layer multilayer perceptron - maps z_t to the predicted temperature change $\Delta\hat{T}(t)$. The surrogate is not autoregressive in temperature: previous temperature predictions are not fed back as inputs, which limits classical forecast-drift effects.

3.3 MARL Climate Mitigation Experiment

We formulate a finite-horizon Markov game played annually from 2016 to 2050 ($H = 35$ time steps) among N agents (representing countries). Each year t, all agents choose actions simultaneously, and these actions determine the emissions and resulting climate outcomes for the next step. The environment is intentionally stylized to isolate the surrogate-integration and policy-fidelity questions, rather than to serve as a complete IAM.

Actions Each agent i's action is a tuple $a_{i,t} = (e_{i,t}, m_{i,t}, l_{i,t}, p_{i,t})$, where $e_{i,t} \in \{0, 0.5, 1.0\}$ is the effort in decarbonizing energy, $m_{i,t}$ is the effort in methane abatement, $l_{i,t}$ is the effort in land-use and agriculture measures, and $p_{i,t} \in \{0, 0.03, 0.08\}$ is the fraction of output invested in adaptation (preventive measures). The three mitigation levers (e, m, l) affect the growth rate of emissions in C via a fixed mapping matrix $M \in \mathbb{R}^{3\times|C|}$ (each action influences each gas in C by a predetermined amount). Given agent i's mitigation effort vector $k_{i,t} = (e_{i,t}, m_{i,t}, l_{i,t})$ the induced growth-rate deviation for gas g is defined as:

$$\tilde{\delta}_{i,g}(t) = \begin{cases} [k_{i,t}\, M]_g & g \in \mathcal{C} \\ 0 & g \in \mathcal{G} \setminus \mathcal{C} \end{cases} \quad (4)$$

and the effective growth factors are then given by:

$$\delta^{\text{eff}}_{i,g}(t) = \delta^{\text{base}}_g(t)\left(1 + \tilde{\delta}_{i,g}(t)\right) \quad (5)$$

The adaptation action $p_{i,t}$ accumulates into a prevention stock P_i that multiplicatively attenuates climate damages in the reward function.

Each agent's emissions are initialized in 2015 as $\bar{E}_i(2015) = S_i \odot E^{\text{base}}(2015)$, where S_i is the vector of the agent's share of baseline emissions (with $\sum_i S_i = \mathbf{1}$). Thereafter, given the agent's effective emission growth factors (after applying its actions), the per-agent emissions are updated each year ($\bar{E}_i(t) = \bar{E}_i(t-1)\odot\delta^{\text{eff}}_i(t)$), so that the global emissions are $E(t) = \sum_{i=1}^{N} \bar{E}_i(t)$. The climate module (either the simulator f_{SCM} or the surrogate f_{NET}) then takes the aggregated emissions and produces the global temperature change $\Delta T(t)$. The surrogate uses only the controllable subset $E_{\mathcal{C}}(t)$ as input, whereas the full simulator uses the entire $E(t)$; both are updated iteratively and carry state forward in time.

Observation Each agent receives the same global observation $O(t)$:

$$O(t) = \big[\Delta T(t-1), \tau(t), \bar{E}_i^{\mathcal{C}}(t-1), D_i^{\mathcal{C}}(t-1), P_i(t-1)\big] \tag{6}$$

where $\tau(t) \in [0,1]$ is a normalized year index, $\bar{E}_i^{\mathcal{C}}(t-1)$ are last-year realized emissions for the controllable gases ($\mathcal{C}$) per agent, $D_i^{\mathcal{C}}(t-1)$ are cumulative deviations from baseline emissions for the controllable gases ($\mathcal{C}$) per agent, and $P_i(t-1) \in [0, P_{\max}]$ are prevention stocks per agent. The two summary quantities are defined as:

$$D_i^{\mathcal{C}}(t) = \sum_{u=2016}^{t} \Big(\bar{E}_i^{\mathcal{C}}(u) - S_i^{\mathcal{C}} \odot E^{\text{base},\mathcal{C}}(u)\Big) \tag{7}$$

$$P_i(t) = \min\big\{P_{\max},\ P_i(t-1)\,\phi + p_{i,t}\big\} \tag{8}$$

where $P_{\max}$ is the maximum effect prevention can have on the climate cost and $\phi \in [0,1]$ is a decay rate on the preventive investments.

Reward At each year t, the reward for agent i is the negative of three types of costs: $C_i^{\mathrm{c}}(t)$ climate disaster costs, $C_i^{\mathrm{k}}(t)$ policy costs, and $C_i^{\mathrm{p}}(t)$ prevention costs:

$$r_i(t) = -\,\eta\Big(C_i^{\mathrm{c}}(t)\ +\ C_i^{\mathrm{k}}(t)\ +\ C_i^{\mathrm{p}}(t)\Big) \tag{9}$$

$$C_i^{\mathrm{c}}(t) = c_i^{\mathrm{c}}\,\psi\,\big(\Delta T(t)\big)^4\,\big(1 - P_i(t)\big) \tag{10}$$

$$C_i^{\mathrm{k}}(t) = \big(c_i^{\mathrm{k}}\big)^{\top}\big(k_{i,t} \odot k_{i,t}\big) \tag{11}$$

$$C_i^{\mathrm{p}}(t) = c_i^{\mathrm{p}} \cdot p_{i,t} \tag{12}$$

where c_i^{c}, $c_i^{\mathrm{k}} = (c_i^{\mathrm{e}}, c_i^{\mathrm{m}}, c_i^{\mathrm{l}})$, and c_i^{p} are the agent-specific climate cost, policy costs, and prevention cost respectively. $P_i(t) \in [0, P_{\max}]$ is the prevention stock, $\psi = 0.003$ is the base climate damage calibration parameter, and $\eta = 10^{-1}$ is the coefficient for reward normalization.

At the terminal step we add a look-ahead cost for near-term climate damage. From the terminal state, $t = 2050$, we roll the climate model forward for $U = 15$ years with baseline emission growth $\delta_g^{\text{base}}(t)$ used to calculate $\bar{E}(t)$ and decaying prevention $P_i(t+u) = \min\{P_{\max},\ P_i(t)\,\phi^u\}$ for $[u, \ldots, U]$:

$$r_i^{\text{term}}(t) = \sum_{u=1}^{U} C_i^{\mathrm{c}}(t+u) \tag{13}$$

$$r_i(t) \leftarrow r_i(t)\ -\ r_i^{\text{term}}(t)\,\eta \tag{14}$$

This terminal look-ahead serves as a pragmatic reward-shaping term to reflect imminent damages in the terminal step, rather than as a general solution to long-horizon credit assignment.

Optimization We train independent recurrent policies using Proximal Policy Optimization (PPO) [Schulman *et al.*, 2017]. Each agent i has parameters θ_i and seeks to maximize its expected discounted return:

$$\theta_i^{\star} = \arg\max_{\theta_i}\ J_i(\theta_i) \tag{15}$$

$$J_i(\theta_i) = \mathbb{E}_{\tau \sim \pi_{\theta_i}}\left[\sum_{t=0}^{H-1} \gamma^t\, r_i(t)\right], \qquad \gamma = 0.999 \tag{16}$$

where $r_i(t)$ is the per-step reward defined in eqs. (9–12). Policies are implemented as LSTM actor-critics trained over complete H=35-year episodes, with the hidden state carried forward through time.

Policy Scenarios We evaluate two versions of the climate-policy game to compare surrogate and simulator performance:

(i) Tractable scenario. $N = 4$ homogeneous agents (identical damage and cost parameters). Only the energy decarbonization lever has a meaningful effect (the other actions are tuned to be prohibitively costly or yield negligible climate impact). This scenario provides strong feedback signals and converges quickly, allowing us to test whether the surrogate-based environment yields the same policies as the original simulator (parameters in Appendix D).

(ii) Intractable scenario. $N = 10$ heterogeneous agents with diverse emission shares, damage sensitivities, and mitigation costs. Multiple mitigation levers produce similar climate outcomes, making it difficult for agents to discern which actions are most effective (resulting in weak learning signals). As a result, this scenario requires vastly more environment interactions to learn good policies. Training to convergence with the full CICERO-SCM model would be prohibitively slow (millions of simulator calls), so we use the surrogate to improve the computational scalability of the climate-transition component in this scenario (parameters in Appendix D).

3.4 Evaluation Criteria

We evaluate the surrogate f_{NET} against f_{SCM}) across four dimensions: accuracy, inference speed, training acceleration, and policy consistency.

Accuracy We compute RMSE and coefficient of determination R^2 between predicted and true temperature increments $\Delta\hat{T}(t)$ and $\Delta T(t)$ on a held-out test set.

Inference Time Both models are wrapped in classes initialized with historical emissions and a step method that appends new emissions $E(t)$, performs preprocessing, and returns $\Delta T(t)$ from inference computation. We measure the mean runtime over 100,000 steps, with CICERO-SCM run on CPU and the RNN surrogate on both CPU and GPU.

MARL Time We compare per environment step time under two otherwise identical environments (scenario (i)) that differ only in the climate backend. CICERO-SCM is executed on CPU whereas the RNN-based surrogate is executed on GPU.

Policy Consistency An ideal surrogate should induce policies indistinguishable from those obtained with the original simulator. Let Π denote the policy class and $J_f(\pi) = \mathbb{E}\Big[\sum_{t=0}^{H} \gamma^t r_t \mid f, \pi\Big]$ denote the expected discounted return under climate engine f and policy $\pi \in \Pi$. We use two criteria to motivate policy consistency:

$$\operatorname{sign}\big[\Delta J_{f_{\text{NET}}}(\pi_1, \pi_2)\big] \approx \operatorname{sign}\big[\Delta J_{f_{\text{SCM}}}(\pi_1, \pi_2)\big] \tag{17}$$

$$\nabla_\theta J_{f_{\text{NET}}}(\pi_\theta) \approx \nabla_\theta J_{f_{\text{SCM}}}(\pi_\theta) \tag{18}$$

where $\Delta J_f(\pi_1, \pi_2) = J_f(\pi_1) - J_f(\pi_2)$ is the reward difference between two policies $\pi_1, \pi_2 \in \Pi$ and $\theta \in \mathcal{N}(\theta^{\star}_{\text{SCM}})$

is the set of parameters in the policy-induced visitation distribution under f_{SCM}. This definition is conceptually related to model-based RL analyses of return and gradient alignment between learned dynamics and true environments [Janner *et al.*, 2021]. In model-based RL terms, eq. (17) relates to preserving policy value ordering, while eq. (18) relates to matching local improvement directions, which depend on surrogate model error along the policy-induced visitation distribution. This is consistent with analyses such as [Shen *et al.*, 2023], which suggest that controlling model error on a policy's visitation distribution controls return discrepancies and helps preserve policy rankings and local improvement directions. However, directly validating eqs. (17)–(18) would require training with f_{SCM}, which is intractable in our large-scale setting.

Hence, we propose a tractable empirical alternative. Let's first define the feasible set of emissions trajectories attainable in our MARL game $\mathcal{K}$ as:

$$\mathcal{K} := \left\{ \bar{E}(\cdot) \;\middle|\; \bar{E}(t), \quad \tilde{\delta}_{\mathcal{G}\setminus\mathcal{C}}(t) = 0, \quad \tilde{\delta}_{\mathcal{C}}(t) \in \mathcal{D}(M, \mathcal{L}) \right\} \tag{19}$$

where the link between $\tilde{\delta}$ and $\bar{E}$ is described in Section 3.3, and $\mathcal{D}$ is the set of controllable growth deviations induced by the lever levels $k_{i,t} \in \mathcal{L}$ and the policy matrix M.

Let $\mathcal{S} \subseteq \mathcal{K}$ be defined as the emission trajectories stored during training using f_{NET} more formally defined as:

$$\mathcal{S} = \{\bar{E}^{(k)}(\cdot)\}_{k=1}^{K} \tag{20}$$

$$\bar{E}^{(k)}(\cdot) = \left(\bar{E}^{(k)}(1), \ldots, \bar{E}^{(k)}(H+U)\right) \tag{21}$$

where $\bar{E}^{(k)}(t) \in \mathbb{R}^{|\mathcal{G}|}$, K denotes the total number of training episodes, H is the episode length, and U is the rollout length. Then let $\bar{\mathcal{S}} \subseteq \mathcal{K}$ denote the set of emission trajectories that would be stored during training using f_{SCM} if that would have been tractable. Trivially, $\mathcal{S} = \bar{\mathcal{S}}$ if $f_{\mathrm{NET}} = f_{\mathrm{SCM}}$ (all else equal).

More generally, if the surrogate uniformly approximates the simulator on $\mathcal{K}$:

$$\sup_{\bar{E}(\cdot)\in\mathcal{K}} \left\| f_{\mathrm{NET}}(\bar{E}(\cdot)) - f_{\mathrm{SCM}}(\bar{E}(\cdot)) \right\|_{\infty} \leq \varepsilon \tag{22}$$

then for sufficiently small ε one expects the sets of policy-induced trajectories to satisfy:

$$\left\| \mathcal{S} - \bar{\mathcal{S}} \right\| \to 0 \quad \text{as} \quad \varepsilon \to 0. \tag{23}$$

While we do not prove this formally, and therefore do not claim policy equivalence, the intuition is that a surrogate that accurately approximates the simulator will have policy-induced emission trajectories $\mathcal{S}$ that are arbitrarily close to those of the true simulator $\bar{\mathcal{S}}$. Moreover, because MARL exploration introduces stochasticity in the agents' policy updates, the sampled trajectories may already overlap substantially with $\bar{\mathcal{S}}$ even for moderate approximation errors ($\varepsilon > 0$).

Therefore, we propose to randomly sample N trajectories from $\mathcal{S}$ (uniform without replacement) and replay them through f_{SCM} to obtain $\Delta T^{\mathrm{SCM}}(\cdot)$. While $\mathcal{S} \neq \bar{\mathcal{S}}$ in general, a sufficiently accurate surrogate implies substantial overlap between $\mathcal{S}$ and $\bar{\mathcal{S}}$, so samples from $\mathcal{S}$ provide a reasonable proxy for the policy-induced emission trajectory distribution under the true simulator $\bar{\mathcal{S}}$. This intuition mirrors the

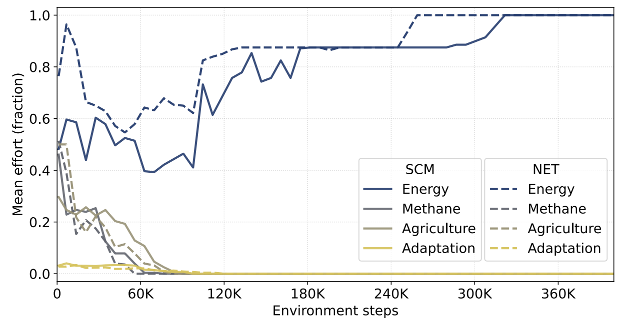

Figure 3: Comparison of learned policies in tractable scenario (i) between CICERO-SCM and the GRU-based surrogate.

transition-occupancy-matching approach presented in [Ma *et al.*, 2023], which learns a dynamics model by matching the distribution of transitions experienced by the current policy in the real environment and in the model. By focusing on policies along the convergence trajectory rather than random states, we evaluate the surrogate in the regions of the state-action space most relevant for converging to the same policies.

Consequently, computing RMSE between $\Delta T^{\mathrm{NET}}(\cdot)$ and $\Delta T^{\mathrm{SCM}}(\cdot)$ for N sampled trajectories in $\mathcal{S}$ provides an empirical proxy for the surrogate's accuracy on policy-induced emission paths. As a proxy for preservation of preference ordering (eq. 17), we also compute Kendall's τ rank-consistency between discounted temperature-based returns defined as:

$$r_k^f = -\sum_t \gamma^t \Delta T_k^f(t), \quad f \in \{f_{\mathrm{SCM}}, f_{\mathrm{NET}}\} \tag{24}$$

where $k = [1, \ldots, N]$. Evaluating only $N \ll |\mathcal{S}|$ trajectories keeps the evaluation tractable when training π^{SCM} to convergence is intractable.

4 Results

4.1 Surrogate Model Performance

The left side of Table 1 summaries the surrogate accuracy and computational efficiency on held-out data. All three surrogates achieve nearly perfect accuracy, with GRU being most accurate (RMSE 3.7×10^{-4} K) and TCN least (RMSE 6.8×10^{-4} K). Given that the training trajectories span 0–2.5K (Figure 2), these errors are very small. LSTM and GRU variants have the fastest one-step inference on GPU (~ 0.0004s), translating to $\sim 1100\times$ faster than CICERO-SCM. This makes embedding climate dynamics in MARL feasible, where simulator use would be intractable. The TCN is slower, requiring 5 layers to reach 10^{-4} accuracy, whereas LSTM and GRU needed only 1 layer. The one-step speedups translate into $> 100\times$ faster per-environment call time during MARL training. Speed-up is not linear with inference time, as policy networks and communication overhead begin to dominate once the surrogate becomes fast. We run 32 environments per runner, further hiding step-function latency. Thus, the comparison is conservative - the speed-up would appear larger without parallelism.

	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}	$442\times$ / $1161\times$	$(5.9 \times 10^{-4}, 0.996)$	$(3.2 \times 10^{-4}, 0.990)$	1.6×10^{-3}	$137\times$
GRU	$(3.7 \times 10^{-4}, 0.99)$	2.3×10^{-3} / 0.4×10^{-3}	$202\times$ / $1161\times$	$(3.9 \times 10^{-4}, 0.996)$	$(2.0 \times 10^{-4}, 0.997)$	1.6×10^{-3}	$137\times$
TCN	$(6.8 \times 10^{-4}, 0.99)$	3.3×10^{-3} / 1.3×10^{-3}	$140\times$ / $357\times$	$(21.1 \times 10^{-4}, 0.994)$	$(10.3 \times 10^{-4}, 0.982)$	4.5×10^{-3}	$49\times$

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.

4.2 Evaluating Policy Consistency

High test-set accuracy does not necessarily mean that the surrogate will induce the same policies as CICERO-SCM. Agents trained with different climate engines may visit different state trajectories, and small prediction errors can compound over long horizons. To assess whether policies learned with f_{NET} remain consistent with those that would emerge under f_{SCM}, we replay $N = 1000$ emission trajectories sampled from the policy-induced distribution visited during training. For each trajectory, we compare the resulting temperature paths from the surrogate and simulator using the RMSE and the Kendall's τ rank-consistency between discounted temperature-based returns. These metrics quantify accuracy along realistic, policy-relevant trajectories and whether the surrogate preserves the preference ordering over policies implied by the simulator.

(i) Tractable Scenario In the homogeneous setting, both surrogate and simulator can be trained to convergence. Table 1 shows that LSTM and GRU maintain low RMSE, and Kendall's τ confirms return rankings are preserved. These results support the policy consistency criteria from eqs. (17–18). Figure 3 shows that, in scenario (i) with only the energy-decarbonization lever activated, the mean lever activation (mean across agents) converges to the same trajectory under both f_{NET} and f_{SCM} (more in Appendix D, Figure A.8).

(ii) Intractable Scenario In the heterogeneous $N = 10$ case, convergence is intractable using the simulator, so we train with f_{NET} for > 1M steps. As shown in Table 1, replayed policy-induced trajectories yield low RMSE and high Kendall's τ, indicating consistent policy ranking under f_{SCM}. A likely reason for the strong replay metrics is that each agent controls a smaller emissions share, leading to smoother, less extreme deviations from SSP2-4.5 and trajectories that remain closer to the surrogate's training distribution. While replay does not provide a formal guarantee, it indicates small discrepancies on the policy-relevant region and preserves the return ordering across policies. Qualitatively, some agents converge to near-max methane abatement, others maintain moderate energy decarbonization, agriculture is used strongly by a subset of agents, while adaptation remains close to zero (Appendix D, Fig. A.8).

5 Limitations and Future Research

The results presented in this paper demonstrate the benefits of using surrogate climate models within climate-economic MARL settings, however, there are limitations and directions for future research.

Refinement of MARL Experiment. More complex experimental setups could incorporate cooperation mechanisms and heterogeneous damage functions, preferably via surrogates of local high-fidelity simulators - an approach our climate collaborators have expressed interest in. Such extensions would primarily modify the climate-impact interface, rather than the MARL loop itself.

Uncertainty-aware Surrogates. We did not perturb the calibration parameters of CICERO-SCM's differential-equation core when training the RNN surrogates. Conditioning the surrogate on these parameters, and potentially predicting distributions rather than point estimates, would propagate structural uncertainty through the MARL loop and enable distributional or risk-sensitive objectives [Bellemare *et al.*, 2017], which are relevant under tipping-point and tail-risk scenarios.

Policy Consistency. While we propose an empirical method to test policy consistency, a formal proof of eqs. (22)–(23) would further substantiate the claim. Our conclusions do not depend on such a proof, and we leave it for future work.

6 Conclusion

In collaboration with a European public-sector environmental agency, we introduced a framework for integrating higher-fidelity climate dynamics into scalable multi-agent reinforcement learning by replacing the climate module with a learned surrogate. We developed an RNN-based emulator of the CICERO-SCM climate model trained on 20,000 multi-gas emission trajectories, achieving global-mean temperature RMSE $< 4 \times 10^{-4}$K and $\sim 1000\times$ faster one-step inference, translating into $> 100\times$ end-to-end MARL training speed-ups.

Using the surrogate within MARL, we converge to the same set of policies in a computationally tractable experiment; when direct validation is infeasible, we propose replaying policy-induced emission trajectories through the simulator as a tractable consistency check. Together, these results remove a major computational barrier to scalable research on cooperative climate-policy design and uncertainty propagation.

Acknowledgments

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

References

[Bellemare *et al.*, 2017] Marc G. Bellemare, Will Dabney, and Rémi Munos. A distributional perspective on reinforcement learning, 2017.

[Biswas *et al.*, 2025] Palok Biswas, Zuzanna Osika, Isidoro Tamassia, Adit Whorra, and et al. Exploring equity of climate policies using multi-agent multi-objective reinforcement learning, 2025.

[Bonan and Doney, 2018] Gordon B. Bonan and Scott C. Doney. Climate, ecosystems, and planetary futures: The challenge to predict life in earth system models. *Science*, 359(6375):eaam8328, 2018.

[Chung *et al.*, 2014] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014.

[Clarke *et al.*, 2022] Leon Clarke, Y.-M. Wei, Ana De La Vega Navarro, Amit Garg, and et al. Energy systems. In P.R. Shukla, Jim Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, and J. Malley, editors, *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, 2022.

[Dorheim *et al.*, 2020] Kalyn Dorheim, Robert Link, Corinne Hartin, Ben Kravitz, and Abigail Snyder. Calibrating simple climate models to individual earth system models: Lessons learned from calibrating hector. *Earth and Space Science*, 7(11):e2019EA000980, 2020. e2019EA000980 10.1029/2019EA000980.

[Emmerling *et al.*, 2016] Johannes Emmerling, Lara A. Reis, Michela Bevione, Loïc Berger, and et al. The witch 2016 model - documentation and implementation of the shared socioeconomic pathways. *SSRN Electronic Journal*, 01 2016.

[Etminan *et al.*, 2016] M. Etminan, G. Myhre, E. J. Highwood, and K. P. Shine. Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing. *Geophysical Research Letters*, 43(24):12,614–12,623, 2016.

[Fricko *et al.*, 2017a] Oliver Fricko, Petr Havlik, Joeri Rogelj, Zbigniew Klimont, and et al. The marker quantification of the shared socioeconomic pathway 2: A middle-of-the-road scenario for the 21st century. *Global Environmental Change*, 42:251–267, 2017.

[Fricko *et al.*, 2017b] Oliver Fricko, Petr Havlik, Joeri Rogelj, Zbigniew Klimont, and et al. The marker quantification of the shared socioeconomic pathway 2: A middle-of-the-road scenario for the 21st century. *Global Environmental Change*, 42:251–267, 2017.

[Fuglestvedt and Berntsen, 1999] Jan S. Fuglestvedt and Terje K. Berntsen. 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), 1999.

[Heitzig *et al.*, 2023] Jobst Heitzig, Jörg Oechssler, Christoph Pröschel, Niranjana Ragavan, and Richie Yat-Long Lo. Improving international climate policy via mutually conditional binding commitments, 2023.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9:1735–1780, 11 1997.

[Intergovernmental Panel on Climate Change (IPCC), 2022] Intergovernmental Panel on Climate Change (IPCC). Summary for policymakers. In H.-O. Portner, D. C. Roberts, E.S. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegria, M. Craig, S. Langsdorf, S. Loschke, V. Moller, and A. Okem, editors, *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, pages 3–33. Cambridge University Press, Cambridge, UK and New York, NY, USA, 2022.

[Janner *et al.*, 2021] Michael Janner, Justin Fu, Marvin Zhang, and Sergey Levine. When to trust your model: Model-based policy optimization, 2021.

[Koasidis *et al.*, 2023] Konstantinos Koasidis, Alexandros Nikas, and Haris Doukas. Why integrated assessment models alone are insufficient to navigate us through the polycrisis. *One Earth*, 6(3):205–209, 2023.

[Kriegler *et al.*, 2017] Elmar Kriegler, Nico Bauer, Alexander Popp, Florian Humpenöder, and et al. Fossil-fueled development (ssp5): An energy and resource intensive scenario for the 21st century. *Global Environmental Change*, 42:297–315, 2017.

[Lea *et al.*, 2016] Colin Lea, Michael D. Flynn, Rene Vidal, Austin Reiter, and Gregory D. Hager. Temporal convolutional networks for action segmentation and detection, 2016.

[Lu and Ricciuto, 2019] Dan Lu and Daniel Ricciuto. Efficient surrogate modeling methods for large-scale earth system models based on machine-learning techniques. *Geoscientific Model Development*, 12(5):1791–1807, 2019.

[Ma *et al.*, 2023] Yecheng Jason Ma, Kausik Sivakumar, Jason Yan, Osbert Bastani, and Dinesh Jayaraman. Learning policy-aware models for model-based reinforcement learning via transition occupancy matching. In Nikolai Matni, Manfred Morari, and George J. Pappas, editors, *Proceedings of The 5th Annual Learning for Dynamics and Control Conference*, volume 211 of *Proceedings of Machine Learning Research*, pages 259–271. PMLR, 15–16 Jun 2023.

[Meinshausen *et al.*, 2011] M. Meinshausen, S. C. B. Raper, and T. M. L. Wigley. 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):1417–1456, 2011.

[Nabuurs *et al.*, 2022] G.-J. Nabuurs, R. Mrabet, A. Abu Hatab, M. Bustamante, and et al. Agriculture, forestry and other land uses (afolu). In P. R. Shukla, Jim Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, and J. Malley, editors, *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, 2022.

[Nicholls *et al.*, 2020] Zebedee R. J. Nicholls, M. Meinshausen, J. Lewis, R. Gieseke, and et al. Reduced complexity model intercomparison project phase 1: introduction and evaluation of global-mean temperature response. *Geoscientific Model Development*, 13(11):5175–5190, 2020.

[Nordhaus and Yang, 1996] William Nordhaus and Zili Yang. A regional dynamic general-equilibrium model of alternative climate-change strategies. *American Economic Review*, 86:741–65, 02 1996.

[Nordhaus, 1992] William D. Nordhaus. An optimal transition path for controlling greenhouse gases. *Science*, 258(5086):1315–1319, 1992.

[Nordhaus, 1993] William D. Nordhaus. Reflections on the economics of climate change. *Journal of Economic Perspectives*, 7(4):11–25, December 1993.

[Nordhaus, 2010] William D. Nordhaus. Economic aspects of global warming in a post-copenhagen environment. *Proceedings of the National Academy of Sciences*, 107(26):11721–11726, 2010.

[Nordhaus, 2014] William Nordhaus. Estimates of the social cost of carbon: Concepts and results from the dice-2013r model and alternative approaches. *Journal of the Association of Environmental and Resource Economists*, 1(1):000, None 2014.

[Nordhaus, 2018] William Nordhaus. Projections and uncertainties about climate change in an era of minimal climate policies. *American Economic Journal: Economic Policy*, 10(3):333–60, August 2018.

[Papoudakis *et al.*, 2021] Georgios Papoudakis, Filippos Christianos, Lukas Schäfer, and Stefano V. Albrecht. Benchmarking multi-agent deep reinforcement learning algorithms in cooperative tasks, 2021.

[Riahi *et al.*, 2017] Keywan Riahi, Detlef P. van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C. O'Neill, Shinichiro Fujimori, and et al. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, 42:153–168, 2017.

[Romero-Prieto *et al.*, 2025] Alejandro Romero-Prieto, Camilla Mathison, and Chris Smith. Review of climate simulation by simple climate models. *EGUsphere*, 2025:1–80, 2025.

[Rudd-Jones *et al.*, 2024] James Rudd-Jones, Fiona Thendean, and María Pérez-Ortiz. Crafting desirable climate trajectories with rl explored socio-environmental simulations, 2024.

[Rudd-Jones *et al.*, 2025] James Rudd-Jones, Mirco Musolesi, and María Pérez-Ortiz. Multi-agent reinforcement learning simulation for environmental policy synthesis, 2025.

[Sandstad *et al.*, 2024] Maria Sandstad, B. Aamaas, A. N. Johansen, M. T. Lund, G. P. Peters, B. H. Samset, B. M. Sanderson, and R. B. Skeie. Cicero simple climate model (cicero-scm v1.1.1) – an improved simple climate model with a parameter calibration tool. *Geoscientific Model Development*, 17(17):6589–6625, 2024.

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

[Schulman *et al.*, 2017] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017.

[Shen *et al.*, 2023] Jian Shen, Hang Lai, Minghuan Liu, Han Zhao, Yong Yu, and Weinan Zhang. Adaptation augmented model-based policy optimization. *Journal of Machine Learning Research*, 24(218):1–35, 2023.

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

[Stehfest *et al.*, 2014] Elke Stehfest, Detlef Vuuren, Tom Kram, Alexander Bouwman, Rob Alkemade, and et al. *Integrated Assessment of Global Environmental Change with IMAGE 3.0. Model description and policy applications*. PBL Publishers, 01 2014.

[Strnad *et al.*, 2019] Felix M. Strnad, Wolfram Barfuss, Jonathan F. Donges, and Jobst Heitzig. Deep reinforcement learning in world-earth system models to discover sustainable management strategies. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(12):123122, 12 2019.

[Weber *et al.*, 2020] Theodore Weber, A. Corotan, B. Hutchinson, B. Kravitz, and R. Link. Technical note: Deep learning for creating surrogate models of precipitation in earth system models. *Atmospheric Chemistry and Physics*, 20(4):2303–2317, 2020.

[Zhang *et al.*, 2022] Tianyu Zhang, Andrew Williams, Soham Phade, Sunil Srinivasa, and et al. Ai for global climate cooperation: Modeling global climate negotiations, agreements, and long-term cooperation in rice-n, 2022.