BOREA-RL

Multi-Objective RL Environment for Climate-Adaptive Boreal Forest Management

Can RL learn forest management policies that **increase carbon sequestration** and **preserve permafrost**?

MOTIVATION

Why Boreal Forest Management Is a Climate Problem

🌍 Carbon Storage

Boreal forests store **30–40% of terrestrial carbon**, much of it locked in climate-vulnerable permafrost soils.

❄️ Permafrost Risk

Permafrost thaw can **reverse carbon storage**, turning boreal ecosystems into a positive climate feedback loop.

The Dual Role of Afforestation

Strategic forest management can serve both climate objectives simultaneously:

- **Remove CO_2** through biomass growth and carbon sequestration
- **Preserve soil carbon** by limiting permafrost thaw through canopy management

This is not just a forestry problem: boreal ecosystems are globally critical, and their mismanagement has planetary consequences.

PROBLEM FRAMING

The Core Trade-Off: Carbon vs. Thaw

The same forest structure can help one objective and hurt the other - making this a **long-horizon, multi-objective** control problem.

Dense Coniferous Stands

- ✅ Better carbon uptake via biomass accumulation
- ❌ Darker canopies increase summer ground energy
- ❌ Snow interactions can accelerate thaw

Deciduous Stands

- ✅ Reduced radiative forcing in some seasons
- ❌ Snow insulation creates competing thaw effects
- ❌ Lower biomass carbon density

Why It's Hard

- Carbon and thaw are coupled through energy balance, snow, species composition, and stand density
- Effects unfold over decades with delayed, noisy feedback

CONTRIBUTION

BoreaRL: A Physically Grounded MORL Environment and Benchmark

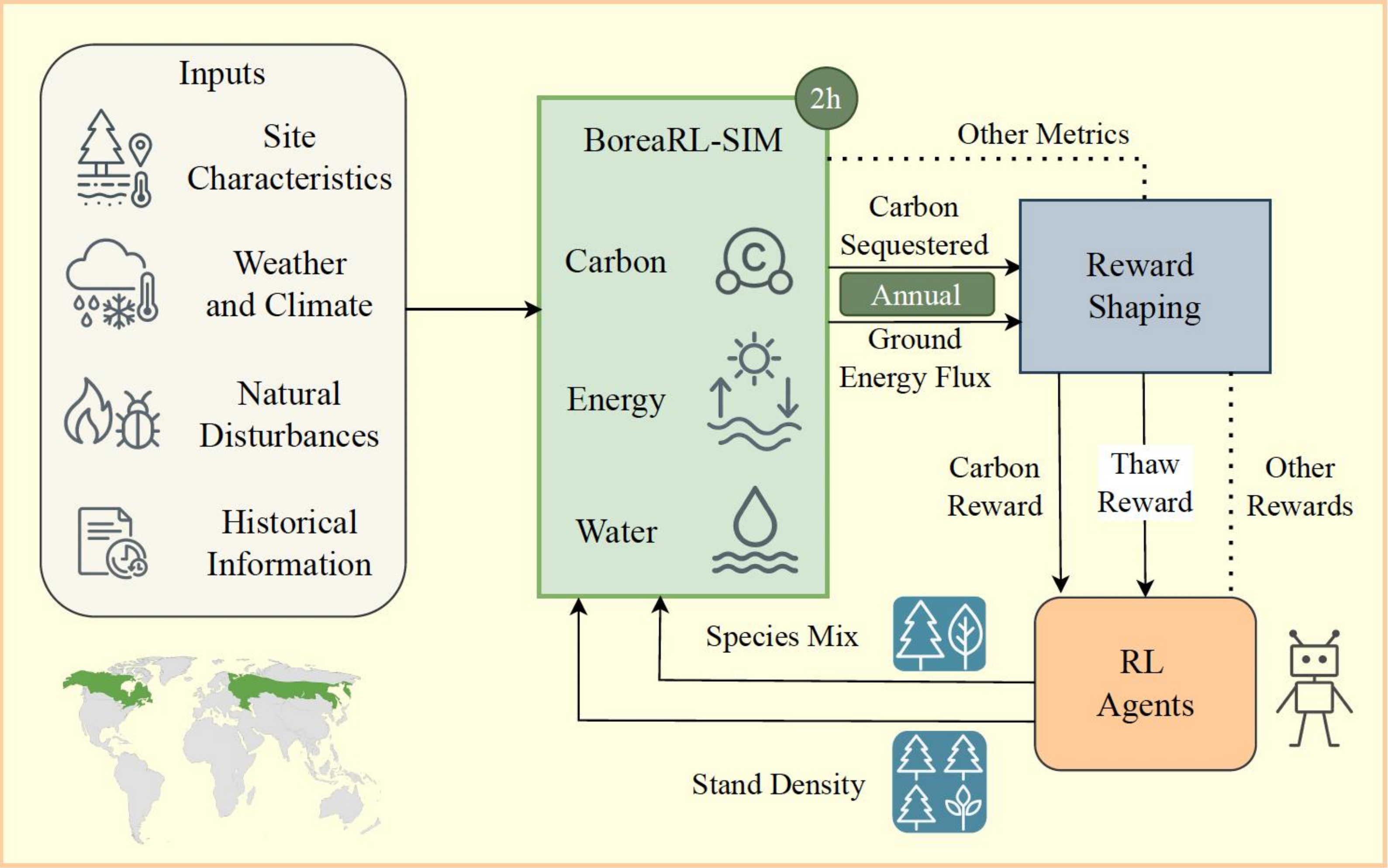
The **first MORL environment** for climate-adaptive boreal forest management.

Process-Based Simulator

Models coupled **energy, water, and carbon** fluxes grounded in ecosystem physics.

MORL Environment Wrapper

Standard multi-objective RL interface supporting both **site-specific** and **generalist** training paradigms.



FORMULATION

How the RL Problem Is Defined

Action Space (Annual Decision)

- **Stand density change:** {−100, −50, 0, +50, +100} stems/ha
- **Target conifer fraction:** {0, 0.25, 0.5, 0.75, 1}

Reward Signals

- **Carbon reward:** biomass-based carbon sequestration
- **Thaw reward:** deep-soil conductive heat flux

Episode Length

50 yearly decisions - long-horizon planning over half a century of management.

Observation space

Ecological state, history, and environmental context.

Thaw is measured via conductive heat flux into deep soil, making it a physically meaningful and challenging signal.

EXPERIMENTS

Benchmark Setting and Baselines

Training Paradigms

Site-Specific

Deterministic, controlled - a single site for reproducible evaluation.

Generalist

Stochastic sites and weather - tests transfer and robustness.

Methods Compared

→ **Variable-λ EUPG**

Preference-conditioned policy gradient method.

$$J_{\text{VarEUPG}}(\theta) = \mathbb{E}_{\lambda\sim\mathcal{D}_\Lambda}\, \mathbb{E}_{\phi\sim\mathcal{D}_{\text{site}}}\, \mathbb{E}_{\tau\sim P^{\phi}_{\pi_\theta}} \left[\sum_{t\geq 0} \gamma^t\, r_t^\lambda\right]$$

→ **PPO Gated**

Preference-conditioned PPO with gated architecture.

→ **Curriculum PPO**

Adaptive episode/site selection.

$$J_{\text{Curriculum}}(\theta,\phi) = \mathbb{E}_{\lambda\sim\mathcal{D}_\Lambda}\, \mathbb{E}_{\phi\sim\mathcal{D}_{\text{site}}}\, \mathbb{E}_{select\sim f_\phi} \left[\mathbb{E}_{\tau\sim P^{\phi}_{\pi_\theta}} \left[\sum_{t\geq 0} \gamma^t\, r_t^\lambda\right] \mid select = 1\right]$$

RESULT 1

Carbon Is Much Easier to Learn Than Thaw

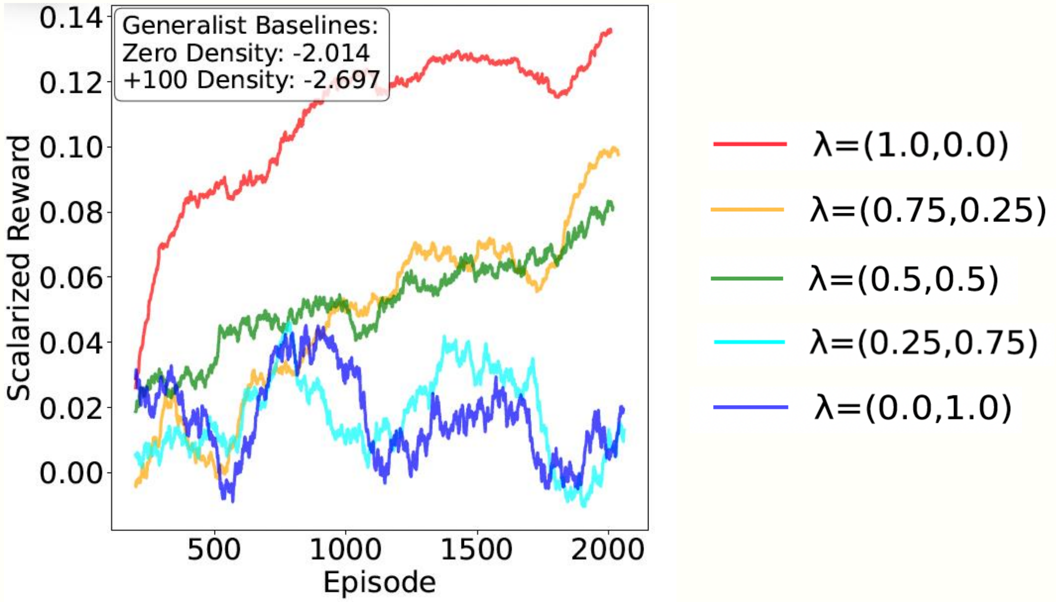
Carbon optimization shows clear gains, while thaw mitigation remains difficult

✅ Carbon Reward

Policies learn quickly in both site-specific and generalist settings. Biomass accumulation gives clear, progressive feedback.

❌ Thaw Reward

Thaw-focused policies show minimal improvement across settings. Signal is delayed, noisy, and risk-asymmetric.



Why the Asymmetry?

Carbon has clear feedback

Biomass accumulates monotonically - the signal is strong and relatively immediate.

Thaw is delayed and noisy

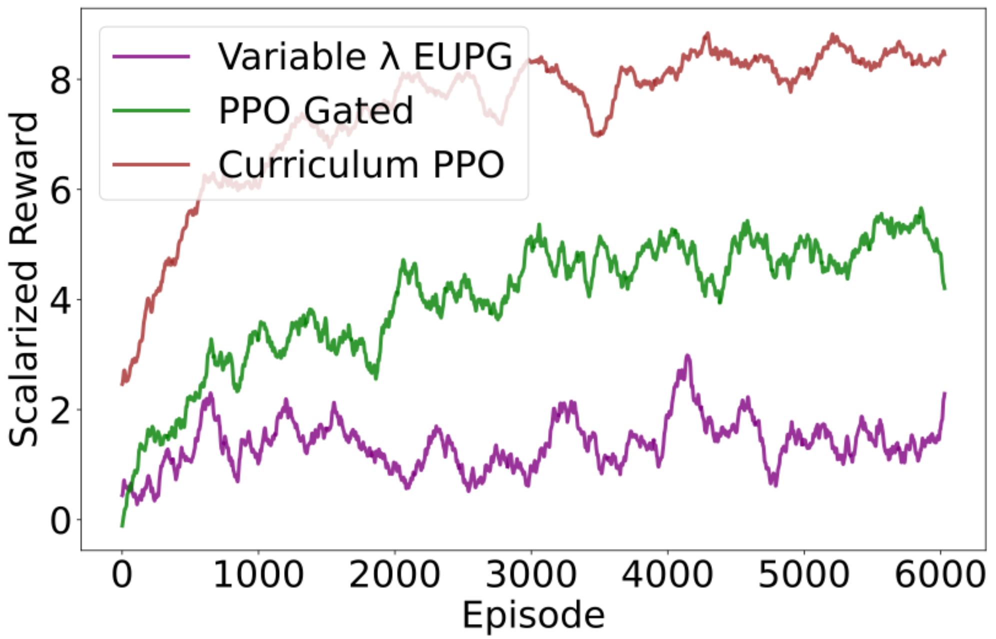
Heat flux into deep soil responds over seasons and years, tied to complex snow-canopy-soil interactions.

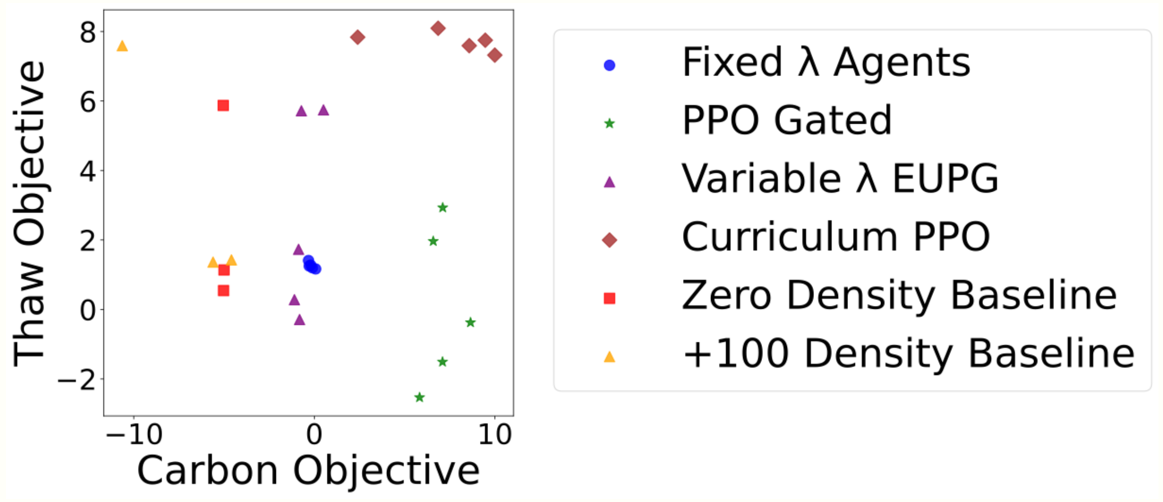
Risk asymmetry

Thaw damage accumulates irreversibly, needing asymmetric reward formulations.

RESULT 2

Simple Site Selection Beats Standard Preference-Conditioned RL





Curriculum PPO performs better likely because adaptive site selection filters out high-variance “trap” episodes, reducing conflicting gradients and making the sparse, noisy thaw signal easier to learn.

RESULT 3

Learned Policies Imply Ecological Strategies

The benchmark reveals interpretable differences in how agents manage stands.

PPO Gated Strategy

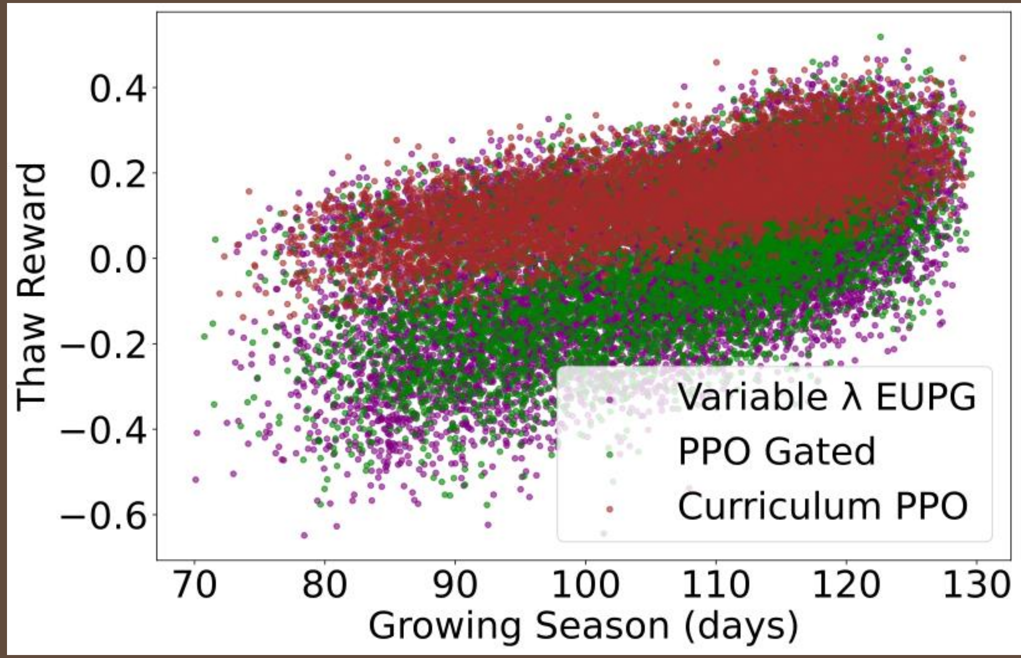
Aggressive high-density, high-conifer "carbon farming"

Maximizes biomass carbon but neglects thaw risk: poor multi-objective trade-off.

Curriculum PPO Strategy

Moderate density + mixed species composition

Balanced decisions yield better thaw protection without large carbon sacrifice.



Key Ecological Insight

Longer growing seasons correlate with better thaw outcomes

TAKEAWAYS

Summary

> Physically grounded climate MORL is feasible, but **robust multi-objective learning remains unsolved.**

New Benchmark: BoreaRL

The first physically grounded MORL environment and benchmark for boreal forest management - open source.

Asymmetric Multi-Objective Problem

Carbon is learnable; permafrost-aware thaw learning is genuinely hard due to delay, noise, and risk asymmetry.

Standard MORL Struggles

Preference-conditioned RL methods underperform in stochastic generalist settings - a clear gap for the community.

Curriculum Is a Strong Path Forward

Adaptive site/episode selection is a surprisingly powerful baseline, pointing to new algorithmic directions.