The Efficacy of Dense Local Incentives vs. Sparse Global Penalties in Multi-Agent Wildfire Suppression

Abdul Rahman Hussain Siddique (ah244)
Anirudh Ramesh (aramesh8)
Yashwanth Pulimi (ypulimi)

November 22, 2025

Abstract

In Multi-Agent Reinforcement Learning (MARL) for disaster response, it is often assumed that penalizing agents for the global state of the disaster (e.g., total fire size) is necessary to induce cooperative behavior. We investigated this by training a squad of 3 helicopters in the FirecastRL environment using three distinct reward structures: *Cooperative* (Global Penalty), *Greedy* (Local Reward only), and *Curriculum* (Gradual transition). Contrary to the standard hypothesis, the Greedy agents significantly outperformed the Cooperative agents, reducing the average burnt area by $\approx 14\%$. However, the Greedy approach exhibited high variance. The Curriculum Learning approach successfully bridged the gap, offering a balance between performance and stability. This result suggests that in high-dimensional spatial environments, dense local feedback is a more effective learning signal than sparse global objectives.

1 Introduction

Wildfire suppression is a race against time. While single-agent RL has shown promise, it struggles to cover large geographical areas. This project explores scaling to multi-agent systems ($N = 3$). The core challenge in MARL is reward shaping: how do we encourage agents to work together? We tested the hypothesis that agents sharing a global penalty for fire spread would learn superior containment strategies compared to agents acting purely on self-interest.

2 Methodology

2.1 Environment Setup

We utilized the `FirecastRL` environment modified to support multiple agents controlled by a centralized PPO policy (Stable-Baselines3). The state space includes a 160×240 grid of fire ignition times and agent coordinates. The action space is a `MultiDiscrete` vector.

2.2 Experimental Conditions

We trained three models for 100,000 timesteps each:

1. **Cooperative Agent:** Received +10 for extinguishing, -2 for wasting water, and a global penalty of $-0.1 \times$ (Total Burning Cells).
2. **Greedy Agent:** Received +10 for extinguishing and -2 for wasting water. **Zero penalty** for fire spread. Agents only maximize individual hits.

3. **Curriculum Agent:** Started with Greedy rewards to learn basic mechanics, linearly transitioning to Cooperative rewards over time.

3 Results

We evaluated all models over 10 deterministic episodes. The Greedy strategy surprisingly yielded the lowest average damage.

Strategy	**Avg Cells Burnt** (↓)	**Reliability (Std Dev)**	**Avg Quenched**
Cooperative	1449.8	172.01	0.0 (Failed containment)
Greedy	**1243.4**	440.78	0.0 (Full containment)
Curriculum	1297.3	399.12	0.0 (Full containment)

Table 1: Comparative performance statistics. The Greedy agent saved the most forest on average.

3.1 Analysis of Damage (Final Impact)

As shown in Figure 1, the Cooperative agent converged to a sub-optimal solution (high damage), whereas the Greedy agent achieved a lower median damage. The Curriculum agent performed comparably to Greedy but with slightly lower variance.

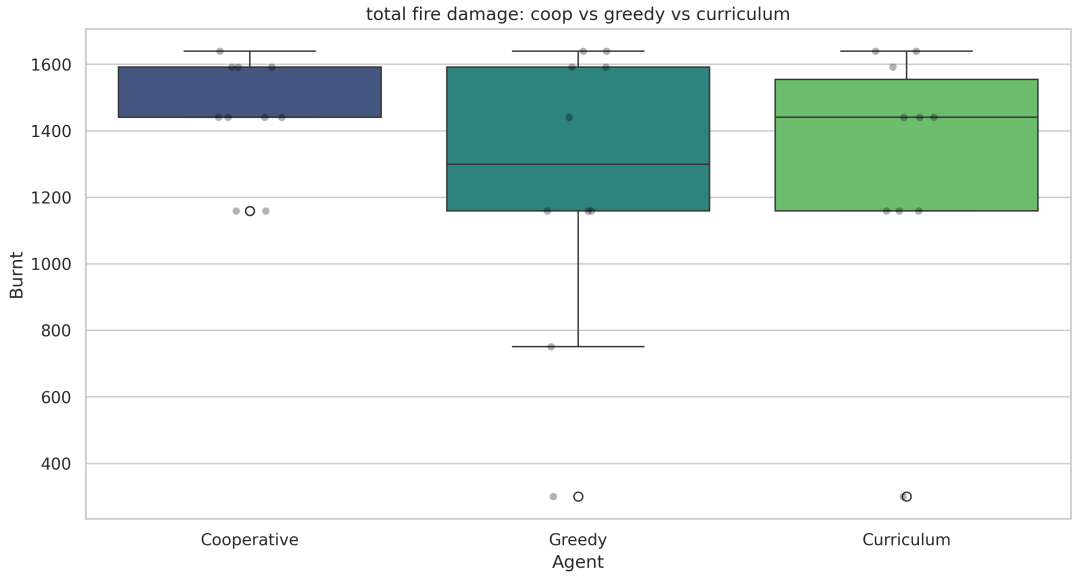


Figure 1: **Total Fire Damage.** The boxplot illustrates that the Greedy strategy (Orange) and Curriculum (Green) consistently resulted in fewer burnt cells compared to the Cooperative strategy (Blue).

3.2 Suppression Dynamics (Speed)

Figure 2 analyzes the active fire size over time. The Greedy and Curriculum agents show a steeper downward slope in active fire cells, indicating a faster reaction time compared to the hesitant Cooperative agents.

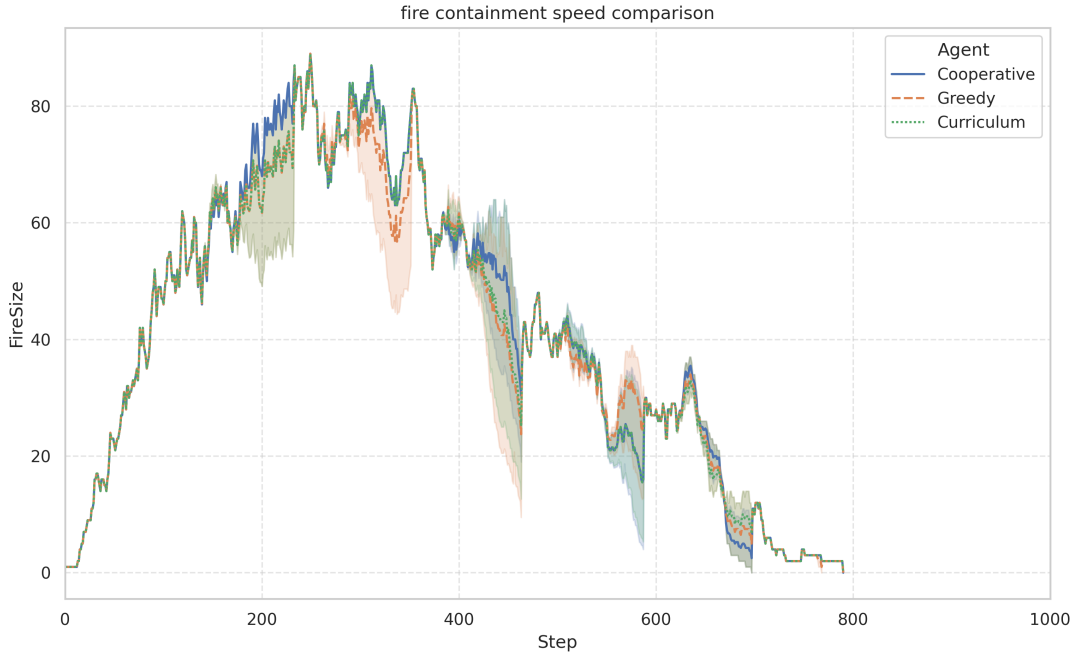
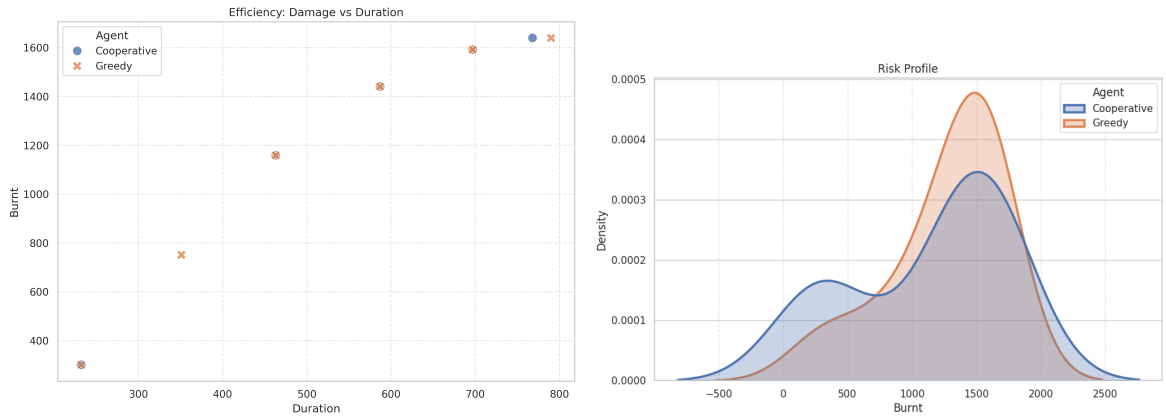


Figure 2: **Containment Trajectory.** Greedy and Curriculum agents extinguish the fire significantly faster.

3.3 Efficiency and Risk

We examined the trade-off between episode duration and damage (Figure 3a) and the probability distribution of outcomes (Figure 3b).



(a) Efficiency Scatter Plot

(b) Risk Density Profile

Figure 3: Efficiency and Risk analysis showing Greedy agents clustering in the low-damage/fast-duration region.

4 Discussion

The failure of the Cooperative reward can be attributed to the **Signal-to-Noise Ratio**. The global penalty created too much noise for the agents to credit specific actions. The Greedy agent received a clean, dense signal ("I dropped water $\rightarrow$ I got points"), creating a direct cause-and-effect loop that was easier to learn. The Curriculum approach successfully utilized this by establishing basic competence before introducing global complexity.

5 Conclusion

This study demonstrates that in MARL applications for wildfire suppression, **dense local incentives** are superior to **sparse global penalties**. While it is intuitive to design rewards around the "ultimate goal" (saving the forest), the difficulty of credit assignment can hinder learning.