Optimizing Rice Palatability in Taiwan through Simulation-Based Reinforcement Learning

Yu-Hsiang Huang [1] Ge-Han Wu [1] Yi-Hsuan Chen [2]

[1]National Tsing-Hua University [2]National Yang Ming Chiao Tung University

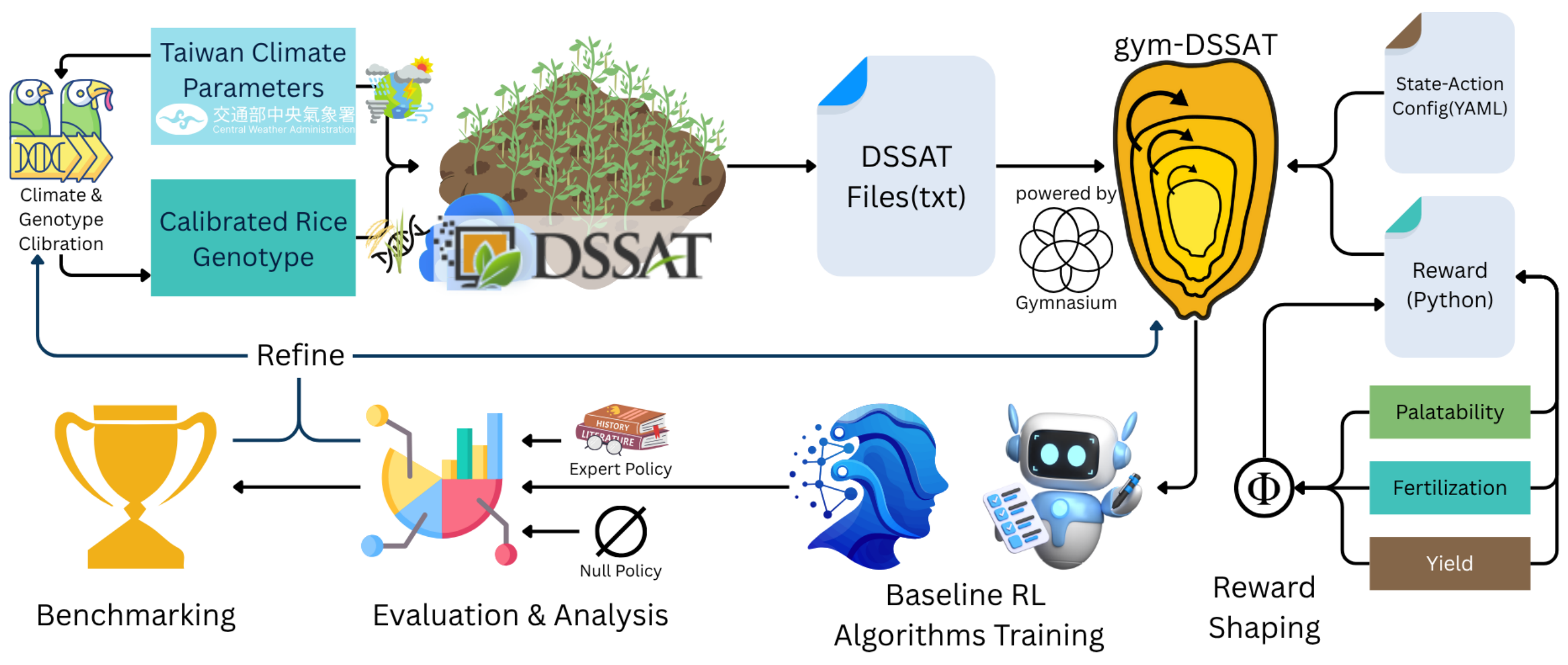
Abstract

This project develops a customized RL environment for rice crop management in Taiwan, aiming to optimize rice palatability while ensuring sustainable yields. The present study investigates the following objectives:

- **Objective 1:** Simulator calibration – Calibrated using local data (e.g., genotype, weather, soil) to accurately reflect Taiwanese rice-growing conditions.
- **Objective 2:** Reward design – Calculated with palatability as the primary objective. Enhanced with reward shaping to provide daily feedback, addressing the sparse reward issue.
- **Objective 3:** Baselines and benchmarking - Train PPO RL models and benchmark them with the null and a Taiwanese expert policy for the palatability optimization problem while ensuring sustainable yields.

Study Methodology

Our environment is modified from the gym-DSSAT framework. The following diagram illustrates the overall workflow and methodology of our study.



Simulator Calibration

Accurately simulating crop growth under Taiwanese conditions requires careful calibration of the DSSAT model. This involves configuring the simulation environment with localized weather data from Taiwan's Central Weather Administration, and incorporating historical yield and palatability data from 2021 as calibration targets. To optimize crop genotype parameters, an evolutionary strategy algorithm is applied, running for 60 generations with a population size of up to 120, thereby improving the model's predictive performance under region-specific conditions. The objective of this calibration is to minimize the discrepancy between simulated outputs and actual field data. The optimization result achieved an absolute error of 2545.97 kg/ha in yield and a 7-point difference in palatability score, indicating a reasonable alignment between the model predictions and observed values.

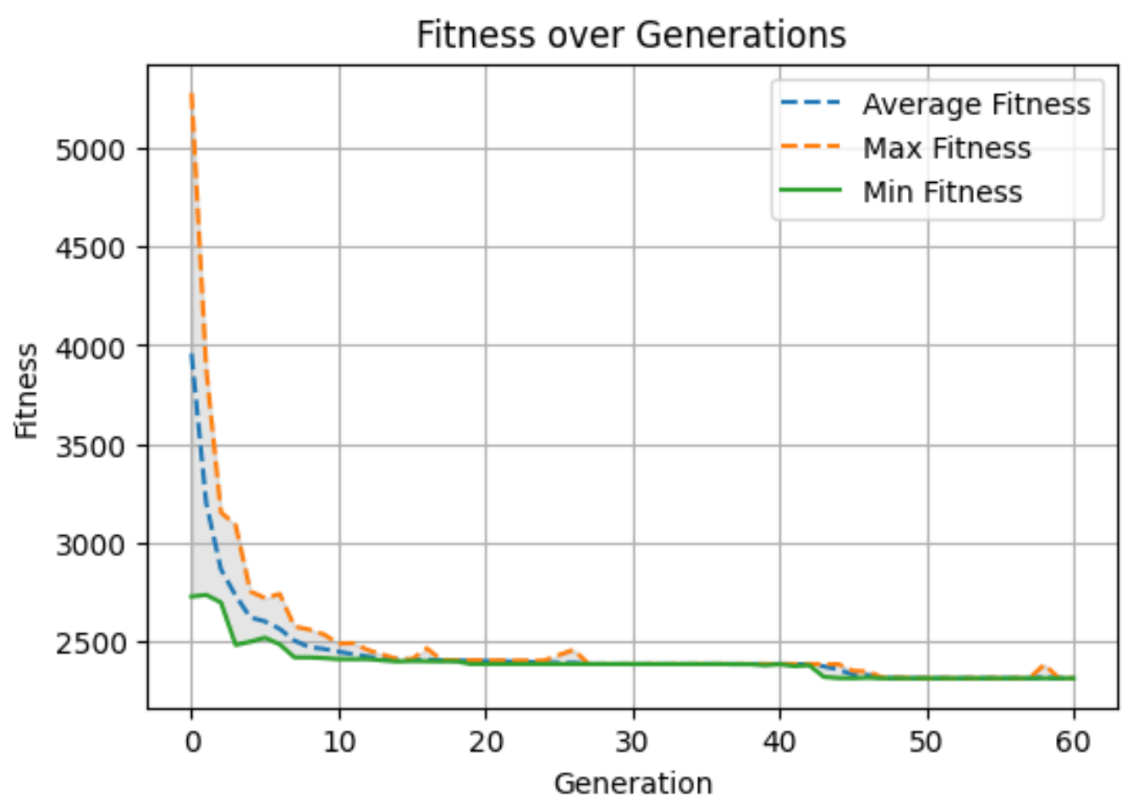


Figure 1. Genotype Calibration using Evolutionary Strategy Algorithm

Reward Design and Shaping

Define a function operator $\Gamma f\left(x,a,b\right) \coloneqq \frac{f(x)-a}{b-a}$. We have the final reward function

$$R_{\text{final}}(s_t) = \begin{cases} \psi_{\text{yield}}(s_t)\cdot\psi_{\text{fert}}(s_t)\cdot\textbf{protein\_score}(s_t), & \text{if } \textbf{yield}(t) > \textbf{yield}_{\text{low}}, \\ & \text{and } \textbf{fert}_{\text{total}}(t) < \textbf{fert}_{\text{high}} \\ -\textbf{reward}\cdot\left[1-\Gamma\,\textbf{yield}\left(t,0,\textbf{yield}_{\text{low}}\right)\right], & \text{if } \textbf{yield}(t) \le \textbf{yield}_{\text{low}} \\ -\textbf{reward}\cdot\Gamma\,\textbf{fert}_{\text{total}}\left(t,\textbf{fert}_{\text{high}},\textbf{fert}_{\text{max}}\right), & \text{otherwise,} \end{cases}$$

where $\psi_{\textbf{yield}}(s_t) = \min\left\{1,\ \Gamma\,\textbf{yield}\left(t,\textbf{yield}_{\text{low}},\textbf{yield}_{\text{thresh}}\right)\right\}$,

$$\psi_{\textbf{fert}}(s_t) = \min\left\{1,\ \left[1-\Gamma\,\textbf{fert}_{\text{total}}\left(t,\textbf{fert}_{\text{thresh}},\textbf{fert}_{\text{high}}\right)\right]\right\},$$

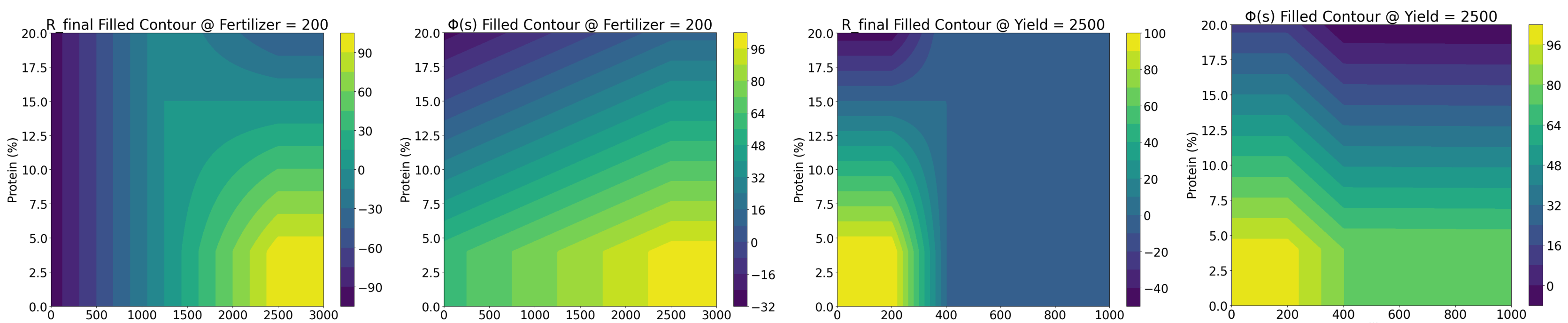
and $\textbf{protein\_score}(s_t) = \textbf{reward}\cdot\min\left\{1,\ \left[1-\Gamma\,\textbf{pr}_{\text{grain}}\left(t,\textbf{pr}_{\text{optimal}},\textbf{pr}_{\text{high}}\right)\right]\right\}$.

We adopted a potential-based reward shaping $R_{\textbf{daily}}(s_t,s_{t+1}) = \Phi(s_{t+1})-\Phi(s_t)$, where

$$\Phi(s_t) = w_{\text{protein}}\cdot\textbf{protein\_score}(s_t) + w_{\text{yield}}\cdot\textbf{yield\_daily}(s_t) + w_{\text{fert}}\cdot\textbf{fert\_daily}(s_t),$$

$$\textbf{yield\_daily}(s_t) = \begin{cases} \textbf{reward}\cdot\psi_{\textbf{yield}}(s_t), & \text{if } \textbf{yield}(t) > \textbf{yield}_{\text{low}} \\ -\textbf{reward}\cdot\left[1-\Gamma\,\textbf{yield}\left(t,0,\textbf{yield}_{\text{low}}\right)\right], & \text{otherwise,} \end{cases}$$

and $\textbf{fert\_daily}(s_t) = \begin{cases} \textbf{reward}\cdot\psi_{\textbf{fert}}(s_t), & \text{if } \textbf{fert}_{\text{total}}(t) < \textbf{fert}_{\text{high}} \\ -\textbf{reward}\cdot\Gamma\,\textbf{fert}_{\text{total}}\left(t,\textbf{fert}_{\text{high}},\textbf{fert}_{\text{max}}\right), & \text{otherwise.} \end{cases}$



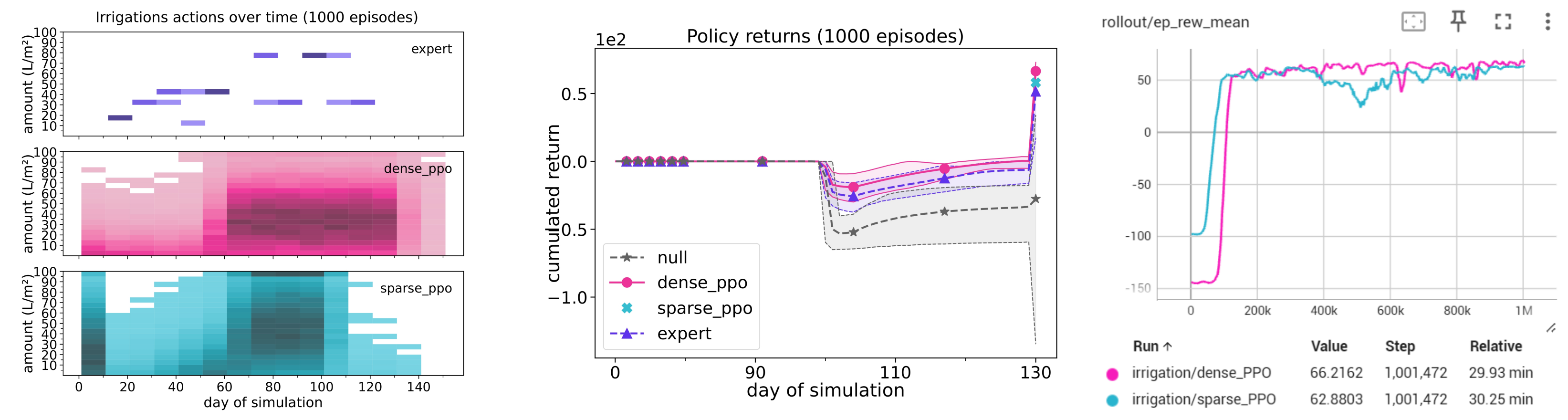
(a) R_{final} contour at $\textbf{fert}_{\text{total}}(t)$=$\textbf{fert}_{\text{thresh}}$ (b) $\Phi(s_t)$ contour at $\textbf{fert}_{\text{total}}(t)$=$\textbf{fert}_{\text{thresh}}$ (c) R_{final} contour at $\textbf{yield}(t)$=$\textbf{yield}_{\text{thresh}}$ (d) $\Phi(s_t)$ contour at $\textbf{yield}(t)$=$\textbf{yield}_{\text{thresh}}$

Baselines and Benchmarking

Policy Descriptions:

1. Null Policy: No action taken.
2. Expert Policy: Uses historical expert data (7 years).
3. PPO (Dense Reward): Trained with daily reward signals.
4. PPO (Sparse Reward): Trained with only final outcome rewards.

Task 1: Irrigation

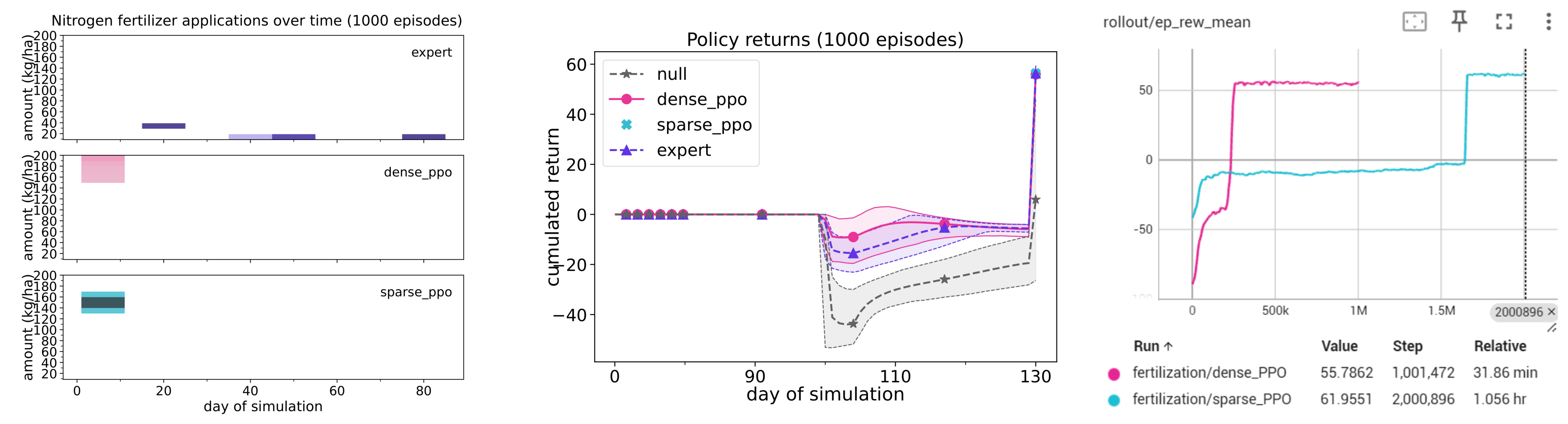


Run ↑	Value	Step	Relative
irrigation/dense_PPO	66.2162	1,001,472	29.93 min
irrigation/sparse_PPO	62.8803	1,001,472	30.25 min

(a) actions (b) reward (c) training curve

Parameter	Treatment			
	Null	Expert	Dense PPO	Sparse PPO
Palatability	64.2(3.5)	70.1(2.4)	**72.1(2.4)**	70(2.4)
Grain Yield (kg/ha)	1790.6(1250.9)	2901.5(521.8)	3391.1(608.4)	**3556.5(646.2)**
Total Irrigation (kg/ha)	0	1015	2677.9(295.2)	2070.7(288.9)
Application Number	0	25	75.4(6.5)	49(5.2)
Water Use Efficiency	N/A	**1.1(1.5)**	0.6(0.6)	0.9(0.7)

Task 2: Fertilization



Run ↑	Value	Step	Relative
fertilization/dense_PPO	55.7862	1,001,472	31.86 min
fertilization/sparse_PPO	61.9551	2,000,896	1.056 hr

(a) actions (b) reward (c) training curve

Parameter	Treatment			
	Null	Expert	Dense PPO	Sparse PPO
Palatability ↑	68.7(2.3)	**69.6(2.3)**	69.3(2.5)	69.5(2.3)
Grain Yield (kg/ha)	1861.7(368.8)	3445.4(413.1)	**4485.6(783.5)**	4100.3(722.7)
Total Fertilization (kg/ha)	0	60(0)	198.4(5.1)	149.7(6.1)
Application Number	0	3	1	1
Nitrogen Use Efficiency	N/A	**26.4(3.9)**	13.2(2.6)	15(3)

Conclusions

- For the DSSAT simulator calibration, we achieved an absolute error of 2545.97 kg/ha in yield and a 7-point difference in palatability score.
- We designed a reasonable and RL-learnable reward function for the palatability optimization problem while ensuring sustainable yields, and adopted a potential-based shaping to mitigate its delayed and sparse characteristics.
- For benchmarking, the Dense PPO yielded the highest palatability score with competitive yield in the irrigation task, and converged faster than the Sparse PPO in the fertilization task, showing the effectiveness of the reward shaping.