Pakistan-Localized Cost-Aware Reinforcement Learning for Simulated Crop Resource Allocation Using CyclesGym

Haider Masood
401636
MS-AI 2k22

Dr. Zuhair Zafar
Dr. Muhammad Moazam Fraz
Dr. Fahad Ahmed Satti
Dr. Syed Imran Ali

PRESENTATION OUTLINE

INTRODUCTION

INTRODUCTION

PAKISTAN AN AGRICULTURAL NATION

- ~22% of GDP.
- Major irrigated area
- Exports share (rice, cotton)
- Employs 36.1% of labor force

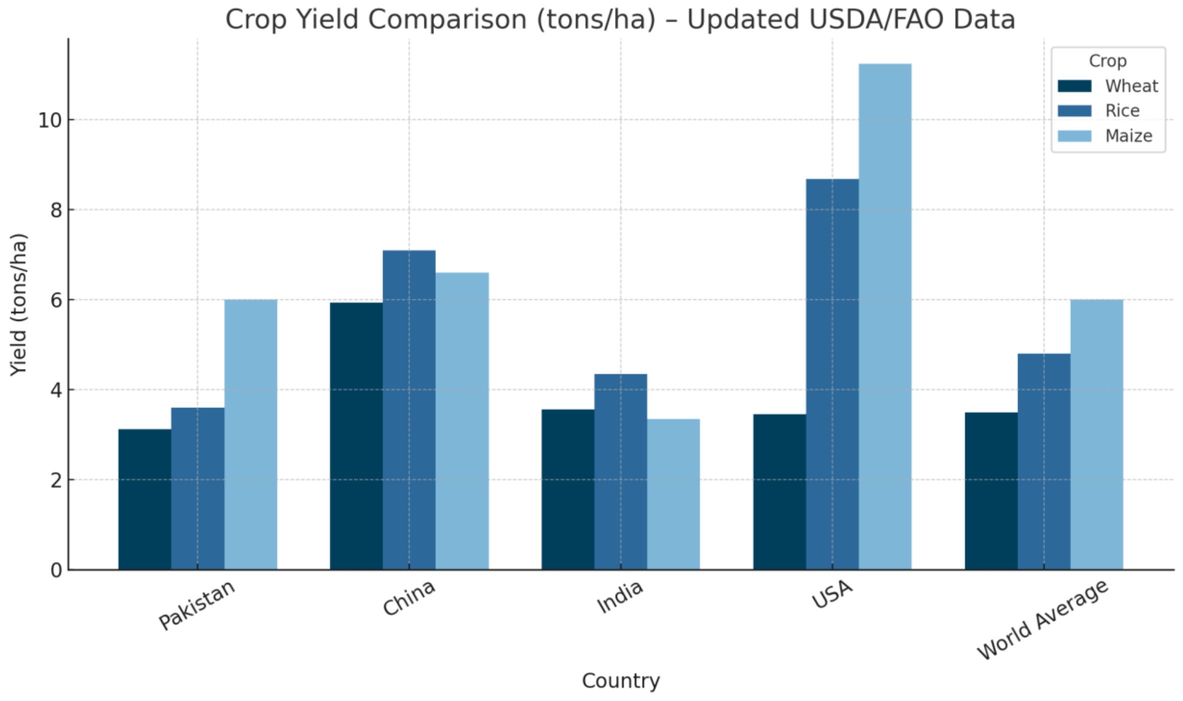
https://www.agripunjab.gov.pk https://tradingeconomics.com/pakistan/employment-in-agriculture-percent-of-total-employment-wb-data.html

INTRODUCTION

Global Agricultural Production Comparison

Pakistan's Agricultural Performance vs. Global Standards

Country	Wheat Yield (tons/ha)	Rice Yield (tons/ha)	Maize Yield (tons/ha)
Pakistan	2.8	3.6	4.9
China	5.4	7	6.3
India	3.5	4	3.2
USA	3.2	8.7	11.1
World Average	3.4	4.6	5.9



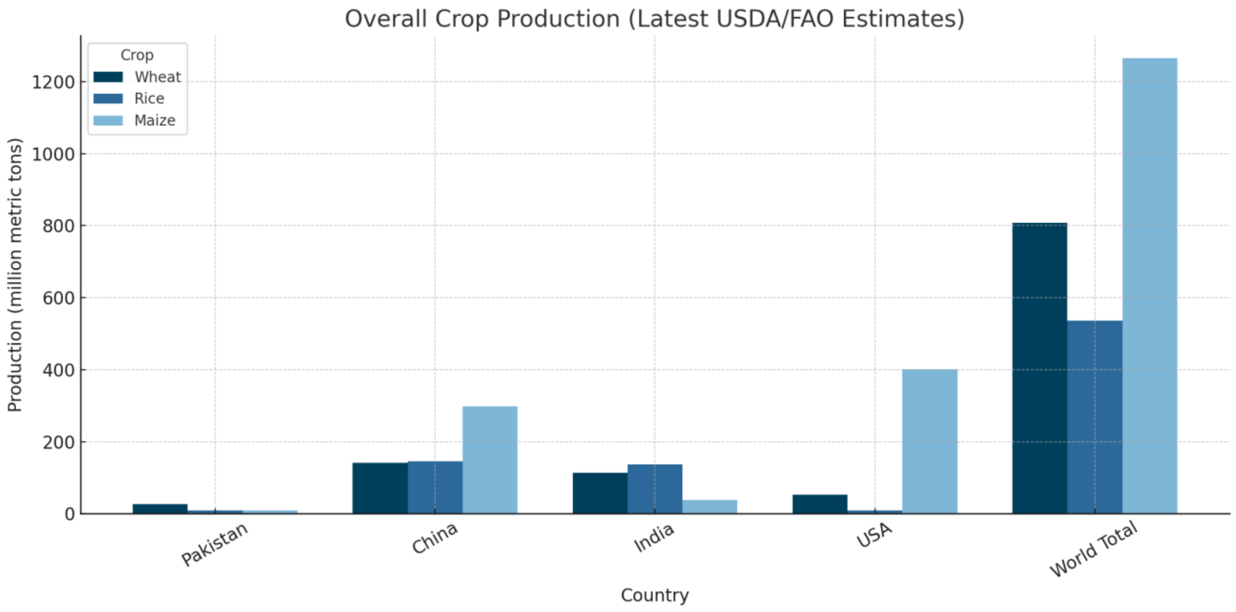
https://ipad.fas.usda.gov/countrysummary/Default.aspx?id=PK&crop=Wheat

INTRODUCTION

Global Agricultural Production Comparison

Pakistan's Agricultural Performance vs. Global Standards

Country	Wheat	Rice (milled-eq.)	Maize
Pakistan	27.5 MMT	9.8 MMT	9.6 MMT
China	141 MMT	146 MMT	298 MMT
India	113.3 MMT	137.8 MMT	37.7 MMT
United States	≈52 MMT (1.921 bn bu) (downloads.usda.library.cornell.edu)	≈10 MMT (222 M cwt) (lsuagcenter.com)	≈402 MMT (15.82 bn bu) (feedandgrain.com)
World total	808.5 MMT (graincentral.com)	535.8 MMT (ers.usda.gov)	1 265 MMT (usda.gov)



https://ipad.fas.usda.gov/countrysummary/Default.aspx?id=PK&crop=Wheat

INTRODUCTION

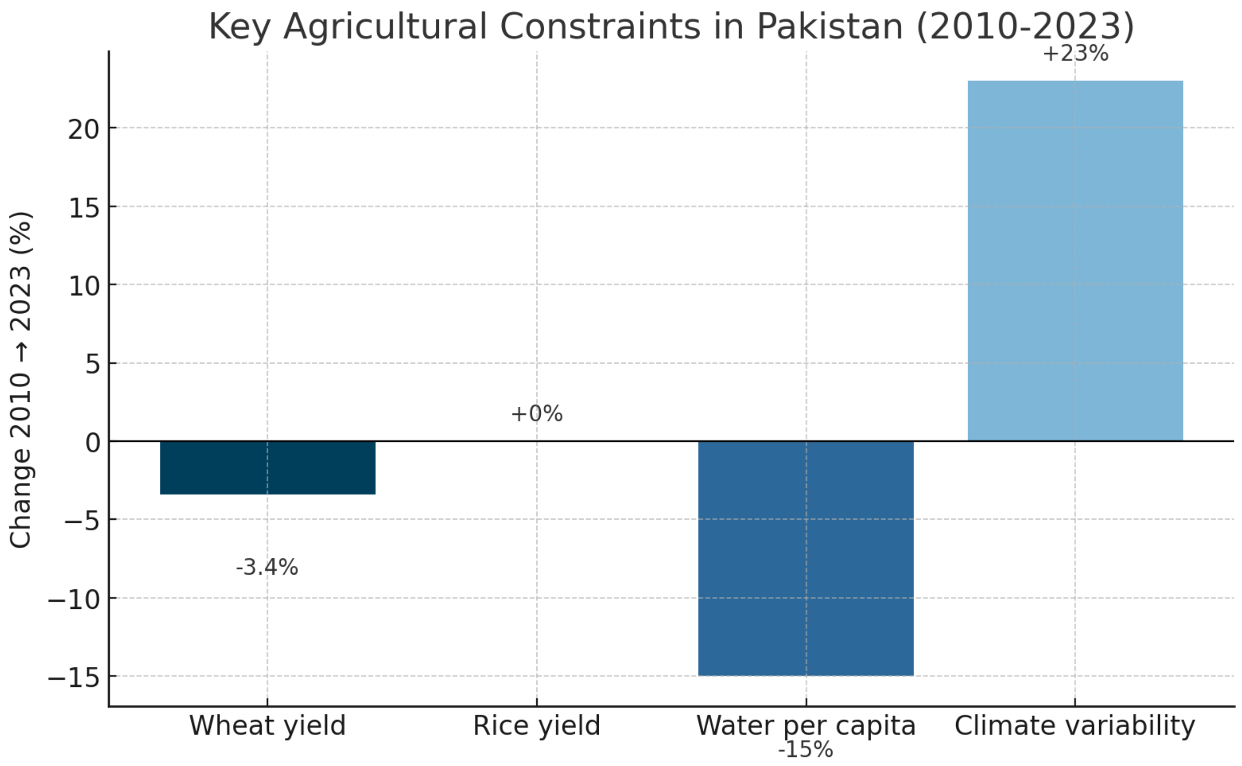
Productivity Headwinds (2010-2023)

Key facts:

- Wheat yield ↓ **3 %** (2.9 → 2.8 t ha^{-1})
- Rice yield **flat** (≈ 3.6 t ha^{-1})
- Per-capita water ↓ **15 %**
- Climate-variability index ↑ **23 %**

Reasons:

- Inefficient fertilizer timing
- Flood-irrigation losses > 40 %
- Low adoption of precision tech
- More heatwaves, floods, droughts



USDA Production Database 2024; FAO AQUASTAT 2023; Pakistan Met Dept. extreme-weather index 2024.

LITERATURE REVIEW

LITERATURE REVIEW

Machine Learning Approaches in agriculture and their limitations

Paper Title	Authors	Strengths	Weaknesses	Dataset Used	Venue & Year
Integrating Random Forest & Crop Modeling Improves Yield of Winter Wheat & Oilseed Rape **[1]**	Dhillon et al.	• Hybrid RF + LUE model lifts R^2 & cuts RMSE vs. RF w/o crop-model inputs. • Uses NDVI + climate ↔ crop growth priors.	• Tested in one European region; transferability unknown. • Needs crop-model calibration.	Sentinel-2 NDVI, ERA5 climate; field yields (Germany)	*Frontiers in Remote Sensing*, 2022
CLIMATES: Mitigating Low Productivity on African Smallholder Farms **[2]**	Tabar et al.	• Clustering + per-cluster ensemble lowers error vs. LSTM/VRNN. • Handles 2.2 k farm time-series with diverse climates.	• Relies on quality cluster labels. • Designed for environmental stress forecasting, not prescriptions.	Environmental time-series (AET, RET, NPP) from 2 200 African farms	AAAI 2022

LITERATURE REVIEW

Machine Learning Approaches in agriculture and their limitations

Paper Title	Authors	Strengths	Weaknesses	Dataset Used	Venue & Year
Optimal Greenhouse Sensor Placement via GBM Grid-Search **[3]**	Uyeh et al.	• Gradient Boosting selects minimal sensors with lowest RMSE. • Psychrometric feature engineering improves predictions.	• Focus on temperature/ humidity only. • Validation limited to two greenhouses.	IoT sensor readings (T, RH) in protected cultivation	*Frontiers in Plant Science*, 2022
Explainable Crop Recommendation with Ensemble ML **[4]**	Shastri et al.	• Compares 10 algorithms; Gradient Boosting highest accuracy. • LIME explains nutrient/weather influence on crop choice.	• Static recommendation (no temporal dynamics). • Tested on Indian soils only.	Soil N-P-K & weather data (India)	*Scientific Reports*, 2025

LITERATURE REVIEW

Deep Learning Approaches in agriculture and their limitations

Paper Title	Authors	Strengths	Weaknesses	Dataset Used	Venue & Year
CropAndWeed Dataset: Multi-Modal Learning for Precision Weeding **[5]**	Steininger et al.	• 8 k aerial RGB images → 112 k instances. • CNN detector benefits from species diversity; supports detection + segmentation.	• Drone imagery only; ground-level viewpoint not tested.	CropAndWeed (74 species, images + meta)	WACV 2023
iGrow: Autonomous Greenhouse Control with Deep RL **[6]**	Cao et al.	• Neural “digital twin” + deep RL → +10.2 % tomato yield, +92.7 % profit vs. experts.	• Domain-specific simulator requires calibration per greenhouse. • RL generalisation across climates untested.	Real tomato greenhouse sensor/control logs	AAAI 2022

LITERATURE REVIEW

Deep Learning Approaches in agriculture and their limitations

Paper Title	Authors	Strengths	Weaknesses	Dataset Used	Venue & Year
Deep Diagnosis: Real-time Apple Leaf Disease Detection **[8]**	Khan et al.	• Lightweight CNN runs in real-time on edge devices; high accuracy on multiple apple diseases.	• Apple-specific; cross-crop transferability untested.	Custom apple leaf image set (lab + orchard)	*Computers & Electronics in Agriculture*, 2 022
Augmentation Invariance & Adaptive Sampling for Aerial Segmentation **[9]**	Tavera et al.	• Consistency loss + adaptive sampling ↑ mIoU on Agriculture-Vision. • Robust to rotations/lighting changes.	• Benefits diminish on ground imagery. • Adds training complexity.	Agriculture-Vision (aerial RGB-NIR)	CVPR Workshops 2022

Machine Learning
&
Deep Learning Techniques
(and their Limitations)

ML & DL TECHNIQUES

Existing AI Approaches in Agriculture

Feature	Machine Learning (Predictive Capabilities)	Deep Learning (Computer Vision & RL)
Core Algorithms	Random Forests, Gradient Boosting	CNNs, Vision-Language Models
Primary Applications	Yield prediction, crop recommendation	Disease detection, weed segmentation etc
Key Strengths	Highly explainable; effectively handles historical time-series data.	Advanced visual recognition; specialized control in controlled environments.

ML & DL TECHNIQUES

Existing AI Approaches in Agriculture

Unified Limitations

While both paradigms offer significant specialized advantages, they share critical bottlenecks when deployed in real-world agricultural environments. Specifically, they suffer from a **lack of robust open-field sequential decision-making.**

Temporal & Dynamic Gaps

Environmental Constraints

Stochastic Variables

REINFORCEMENT LEARNING

REINFORCEMENT LEARNING

Why RL - not CSP or Classical Optimisation?

Property	Classical optimisation / CSP	Reinforcement learning (ours)
Problem form	One-shot feasible/optimal plan for a fixed, known scenario	Sequential closed-loop policy a = pi(s) over the season
Model needed	Explicit, tractable objective + constraints	Learns from the black-box CYCLES simulator (no closed form)
Uncertainty	Re-solve per weather outcome (open-loop plan)	One policy adapts online to stochastic monsoon weather
Action space	Must be enumerated by hand; 1,331 NPK x 53 wks explodes	Native large multi-discrete action space (11x11x11)
Output	A single fixed schedule	A reusable policy that generalises to unseen weather

Bottom line: We must supply an action space and let the agent learn the decision rule under uncertainty - that is exactly what RL provides. MPC is the closest classical alternative, but it must re-solve online every step and assumes a model of the field we do not have.

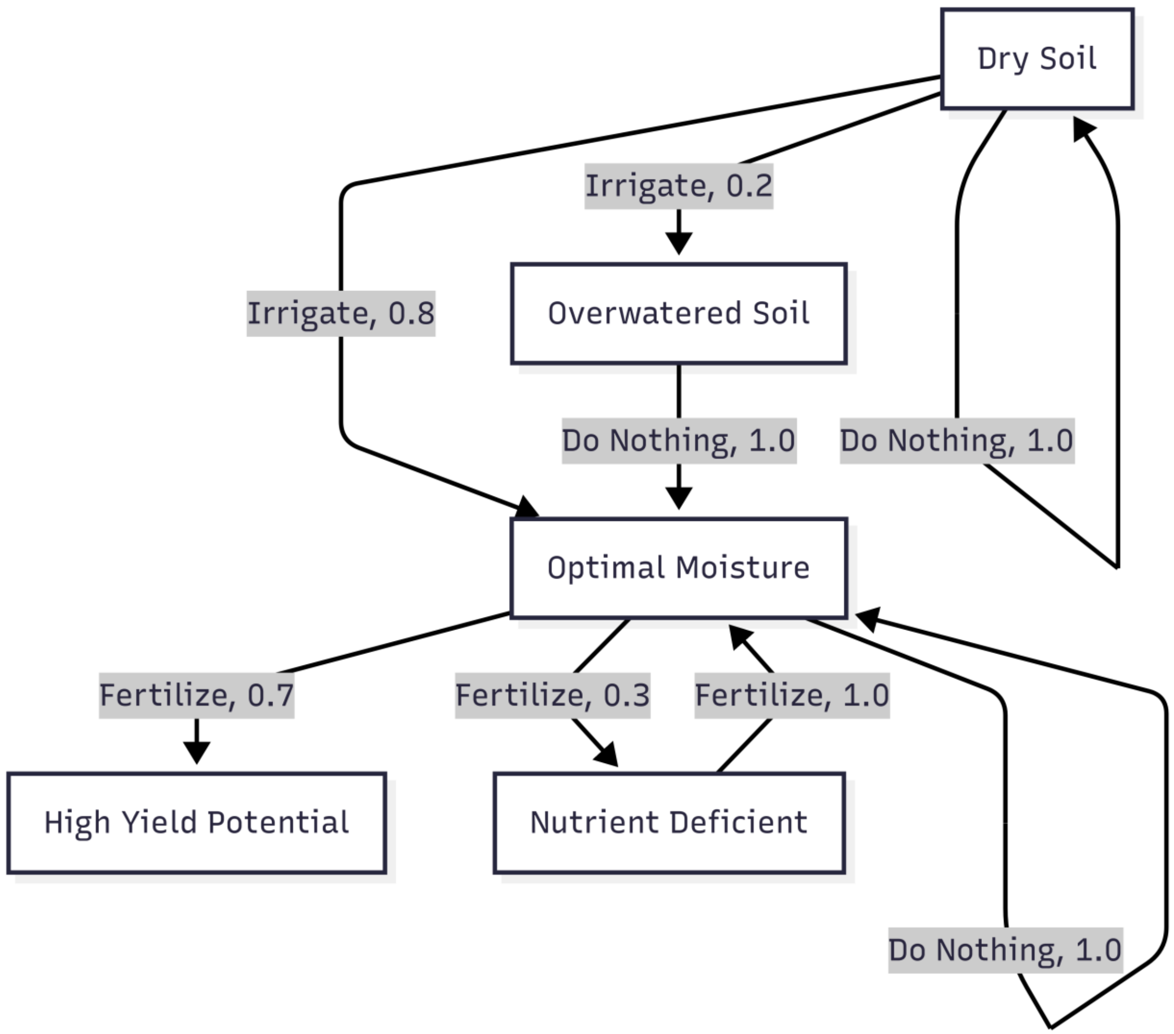
REINFORCEMENT LEARNING

Why Reinforcement Learning for Agriculture ?

- Interactive learning - agent explores the crop/soil/climate simulator through trial-and-error and continuously improves its policy.

- Sequential decision-making (MDP) - models season-long input choices as a Markov Decision Process to maximize long-term reward (crop revenue - input cost).

- Adaptive policy learning - learns state-conditioned fertilizer/crop-planning actions under weather uncertainty; in this thesis, irrigation remains fixed.

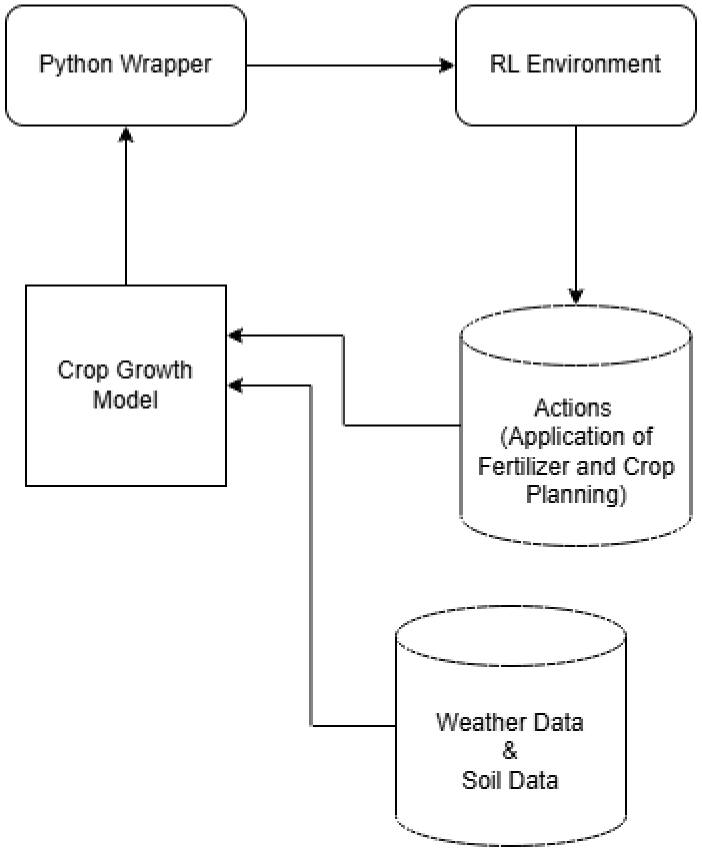


REINFORCEMENT LEARNING

Crop Growth Models and the RL learning cycle

- **Crop Growth Models (CGMs)**
- **Mathematical simulation of plant physiology**
- **Deterministic or Stochastic process simulation**
- **Fundamental tool for large-scale agricultural planning**

- **BENEFITS OF CGMs:**
- **Safe, low-cost testing of management policies**
- **Long-term decision optimization (like Multi-year planning)**
- **Generating high-fidelity simulated evidence before field-adjacent validation**



LITERATURE REVIEW

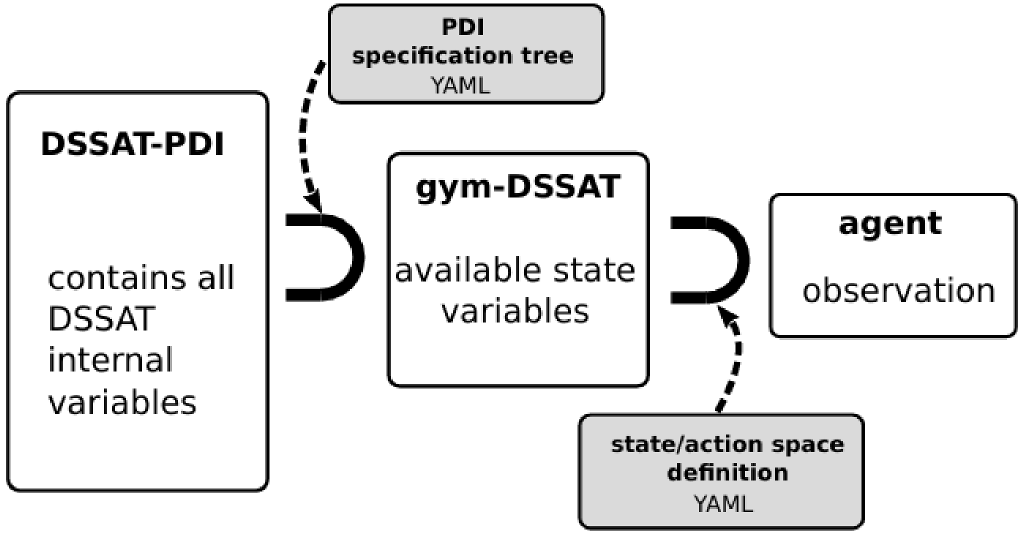
RL PAPER 1 : CropGym: a Reinforcement Learning Environment for Crop Management (Environmental Data Science by Cambridge University Press)

- **Model Foundation:** LINTUL-3 winter wheat module (light-use efficiency based)
- State Representation (18-dimensional)
  - Development stage, Leaf area index, Total biomass (kg/ha), Available soil nitrogen
  - Environmental Context : Temperature, Rainfall, Solar radiation, 5-day aggregated summaries
  - Management Tracking : Recent Fertilization/ Irrigation Flags
  - Action Space : Agent applies nitrogen fertilizer at 5-day intervals based on observed state
- Optimization Framework
  - $r_t = \frac{m_{so,t} - m_{so,t-1}}{m^*_{so,t} - m^*_{so,t-1}} - \beta m_{fert,t}$
  - The reward balances crop yield improvement against fertilizer cost, where β controls the economic-environmental trade-off (β=1 for profit, β>1 for sustainability).

LITERATURE REVIEW

RL PAPER 2 : Gym-DSSAT: a crop model turned into a Reinforcement Learning environment (Research Report N° 9460 by Inria)

- **Crop Model:** DSSAT (Decision Support System for Agrotechnology Transfer)
- **Implementation:** Fortran-Python coupling via PDI
- **Crops:** Maize (extensible to 41+ crops)
- **Episode:** Full growing season (~160 days)
  Decisions: Daily time steps
  Weather: WGEN stochastic generator
  Termination: Crop maturity or early failure
- **State Space (13 variables)**
- Growth stages, biomass, nitrogen stress, leaf area, weather, cumulative inputs



Reward Functions
Fertilization:

$$r(t) = \mathrm{trnu}(t, t+1) - 0.5 \times \mathrm{anfer}(t)$$

Irrigation:

$$r(t) = \Delta\mathrm{biomass}(t) - \mathrm{penalty} \times \mathrm{irrigation}(t)$$

LITERATURE REVIEW

Why Long-Term Planning in Agriculture Needs AI

Modern agriculture faces a multi-objective challenge:

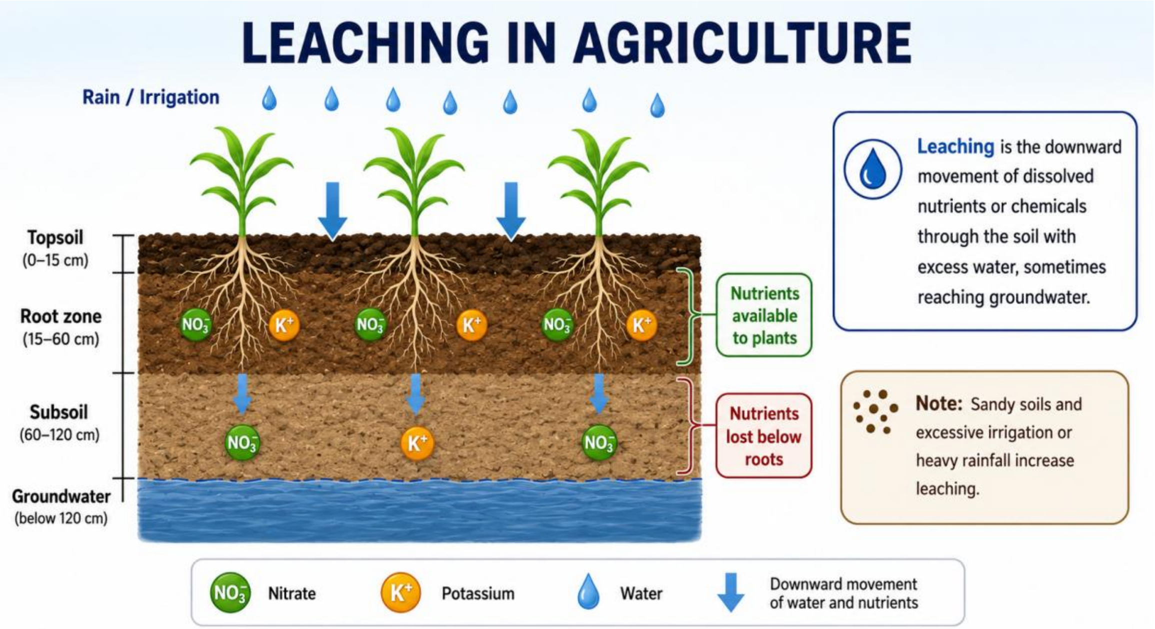
- Maximize profit while minimizing environmental harm and adapting to weather variability and soil degradation.

Farmers often make single-season decisions, but in reality, agriculture is interdependent across years:

- Fertilizer applied this year affects soil nitrogen next year
- Crop choices affect pests, diseases, and water balance
- Unplanned fertilizer use causes waste and pollution

What is Leaching?

- When **too much nitrogen** fertilizer is applied, the excess **leaches** — it dissolves in rainwater and **moves beyond the root zone**



LITERATURE REVIEW

RL PAPER 3 : Long-Term Crop Management Strategies with CyclesGym
(NeurIPS 2022)

	Year 1		Year 2		Reduced Leaching	Reduced Emissions	Profit
Environment							
Management Strategies	Maize	Fertilization	Maize	Fertilization	✗	✗	✓
	Soy		Soy		✓	✓	✗
	Soy		Maize	Low Fertilization	✓	✓	✓

Strategy	Profit	Environmental impact	Why / mechanism
Monoculture **maize** + **non-organic fertilization**	**High ↑**	**High ↑**	Heavy fertilizer → big yields, but **N leaching + GHG emissions** rise
Soybean focus with **low fertilization**	**Lower ↓**	**Low ↓**	Lower yields from soy; minimal external N input
Maize–soybean rotation, leverages **soy N-fixation**	**High ↑**	**Medium–Low ↘**	Soy adds N biologically → **less fertilizer** for maize, decent yields

LITERATURE REVIEW

RL PAPER 3 : Long-Term Crop Management Strategies with CyclesGym
(NeurIPS 2022)

Multi-Year Planning:

$$V^{\pi}(s) = \mathbb{E}^{\pi}\left[\sum_{t=0}^{T} \gamma^{t} r_{t} \middle| s_0 = s\right]$$

where T spans multiple growing seasons

Crop Rotation Benefits:

- Nitrogen fixation modeling:

$$N_{available,t+1} = N_{soil,t} + N_{fixed,t} + N_{fertilizer,t}$$

- Soil health improvements

Constraint Integration:

- Environmental limits: $\text{Leaching}_t \leq \text{threshold}$
- Economic constraints: $\sum_t \text{costs}_t \leq \text{budget}$
- Regulatory compliance modeling

Complex Action Spaces:

- Timing decisions: $a_t = (crop\_type, plant\_date, fertilizer\_amount, timing)$
- Multi-dimensional optimization
- Seasonal planning integration

Available CGM Environments

Characteristic	CropGym	Gym-DSSAT	CyclesGym
Crop Model	LINTUL-3	DSSAT	Cycles
Fertilization Interval	Fixed 5-day	Daily	Continuous
Irrigation Control	Not specified	Daily	Not specified
Strategic Planning	Not specified	Inability to handle multi-year	Supports continuous multi-year
Crop Rotation	Not specified	Not specified	Demonstrates performance benefits
Data Scope	European winter-wheat	Not specified	USA-centric

WHY CYCLES GYM ?

Characteristic	CropGym	Gym-DSSAT	CyclesGym
Crop Model	LINTUL-3	DSSAT	Cycles
Fertilization Interval	Fixed 5-day	Daily	Continuous
Irrigation Control	Not specified	Daily	Not specified
Strategic Planning	Not specified	Inability to handle multi-year	Supports continuous multi-year
Crop Rotation	Not specified	Not specified	Demonstrates performance benefits
Data Scope	European winter-wheat	Not specified	USA-centric

Research Problem

RESEARCH PROBLEM

Problem statement

Problem Statement

"Existing reinforcement learning-based agricultural management systems are developed for North American and European contexts. They cannot be directly applied to Pakistani agriculture due to vastly different climates, crop varieties, soil, crop calendars and economic constraints. We need to address the need for a Pakistan-adapted, cost-aware agricultural RL platform that can study cost-aware crop-planning and NPK fertilization decisions in simulation while producing reproducible experimental evidence."

Solution Statement

SOLUTION STATEMENT

This thesis develops **CyclesGym-PK**, a Pakistan-adapted reinforcement learning framework built on the CYCLES crop simulator to simulate cost-aware agricultural decision-making.

The solution integrates Pakistan-specific **weather, soil, crop-calendar, crop-price, and fertilizer-price data**, extends fertilizer control from nitrogen-only to full **NPK management**, and implements RL environments for **weekly fertilization**, **annual crop planning**, and **hierarchical crop planning plus fertilization**.

The framework trains and evaluates RL agents under fixed and random weather conditions, using a reward function based on: **Crop Revenue (NPK Fertilizer Cost)**

Research Objectives

RESEARCH OBJECTIVES

Research Objectives

Research Objectives:

- Adapt CyclesGym for Pakistan using localized weather, soil, crop-calendar, and price data.
- Extend the fertilizer control system from nitrogen-only decisions to NPK-based nutrient management.
- Design a cost-aware reward function based on crop revenue minus nutrient costs.
- Implement hierarchical crop-planning plus fertilization.
- Train and evaluate PPO, A2C, and selected DQN configurations under fixed and random weather settings.

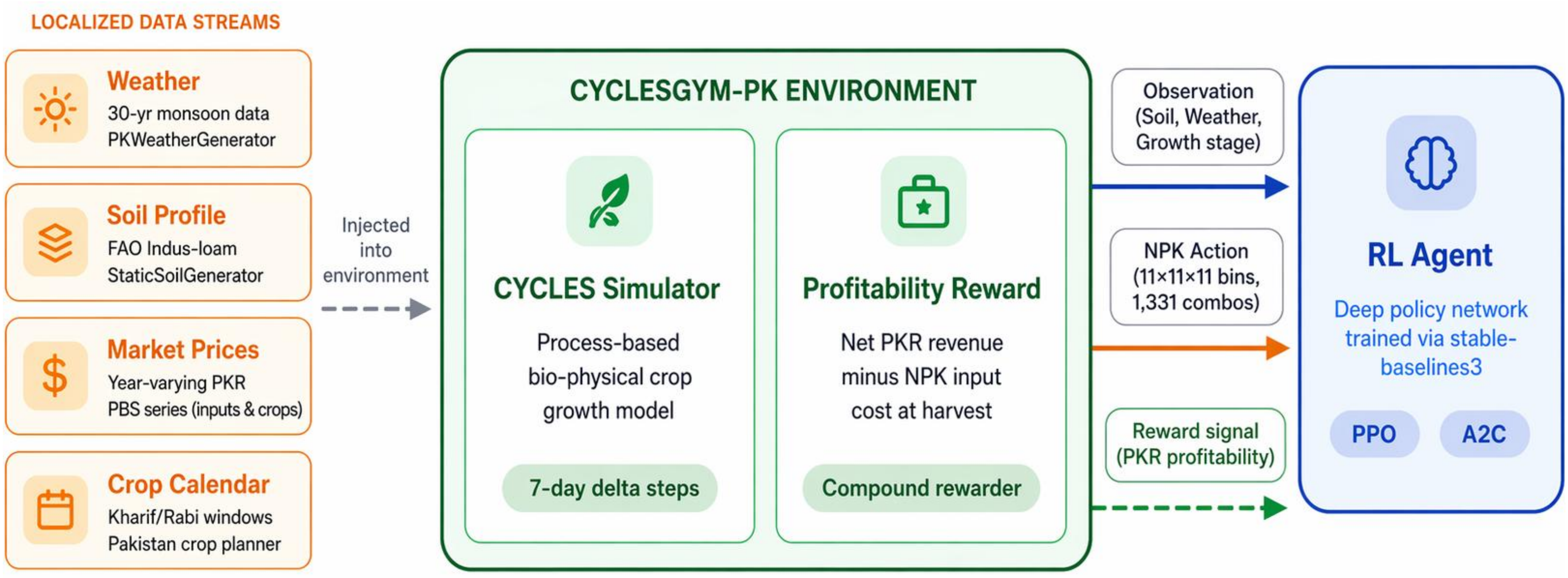
NUST
Defining Futures

MACHVIS
MACHINE VISION
& INTELLIGENT SYSTEMS LAB

Methodology

METHODOLOGY

Architecture of CyclesGym-PK



METHODOLOGY

RL Intuition: How Does the Training Work ?

The Reinforcement Learning Loop in Action

- Observe **(State):** The "AI Farmer" observes the soil NPK levels, plant growth stage, and current weather.
- Act **(Decision):** Agent chooses a weekly fertilizer dose (e.g., 20kg N, 10kg P, 0kg K).
- Simulate **(Environment):** The CYCLES bio-physical model simulates growth and environmental leaching for 7 days.
- Reward **(Feedback):** At harvest, the Agent receives a reward:

(Revenue in PKR) – (Fertilizer Costs)

- Positive reward strengthens the policy
- Negative reward weakens it.

METHODOLOGY

PPO Intuition

The idea with Proximal Policy Optimization (PPO) is that we want to improve the training stability of the policy by limiting the change you make to the policy at each training epoch: **we want to avoid having too large policy updates.**

For two reasons:

- We know empirically that smaller policy updates during training are **more likely to converge to an optimal solution.**
- A too big step in a policy update can result in falling “off the cliff” (getting a bad policy) **and having a long time or even no possibility to recover.**

NUST
Defining Futures

METHODOLOGY

MACHVIS
MACHINE VISION
& INTELLIGENT SYSTEMS LAB

PPO Intuition

$\pi_\theta(a_t|s_t)$ $\pi_{\theta_k}(a_t|s_t)$ $\underset{\theta}{\text{maximize}} \quad \hat{\mathbb{E}}_t\left[\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)}\hat{A}_t\right]$

$\pi_{\theta_{k+1}}(a_t|s_t) \leftarrow \pi_\theta(a_t|s_t)$ $r_t(\theta) = \pi_\theta(a_t|s_t)/\pi_{\theta_k}(a_t|s_t)$

$$\mathcal{L}^{CLIP}_{\theta_k}(\theta) = \underset{\tau\sim\pi_k}{\mathrm{E}}\left[\sum_{t=0}^{T}\left[\min(r_t(\theta)\hat{A}^{\pi_k}_t, \text{clip}\,(r_t(\theta), 1-\epsilon, 1+\epsilon)\,\hat{A}^{\pi_k}_t)\right]\right]$$

Input: initial policy parameters θ_0, clipping threshold ϵ
for $k = 0, 1, 2, \ldots$ **do**
Collect set of partial trajectories $\mathcal{D}_k$ on policy $\pi_k = \pi(\theta_k)$
Estimate advantages $\hat{A}^{\pi_k}_t$ using any advantage estimation algorithm
Compute policy update

$$\theta_{k+1} = \arg\max_\theta \mathcal{L}^{CLIP}_{\theta_k}(\theta)$$

by taking K steps of minibatch SGD (via Adam), where

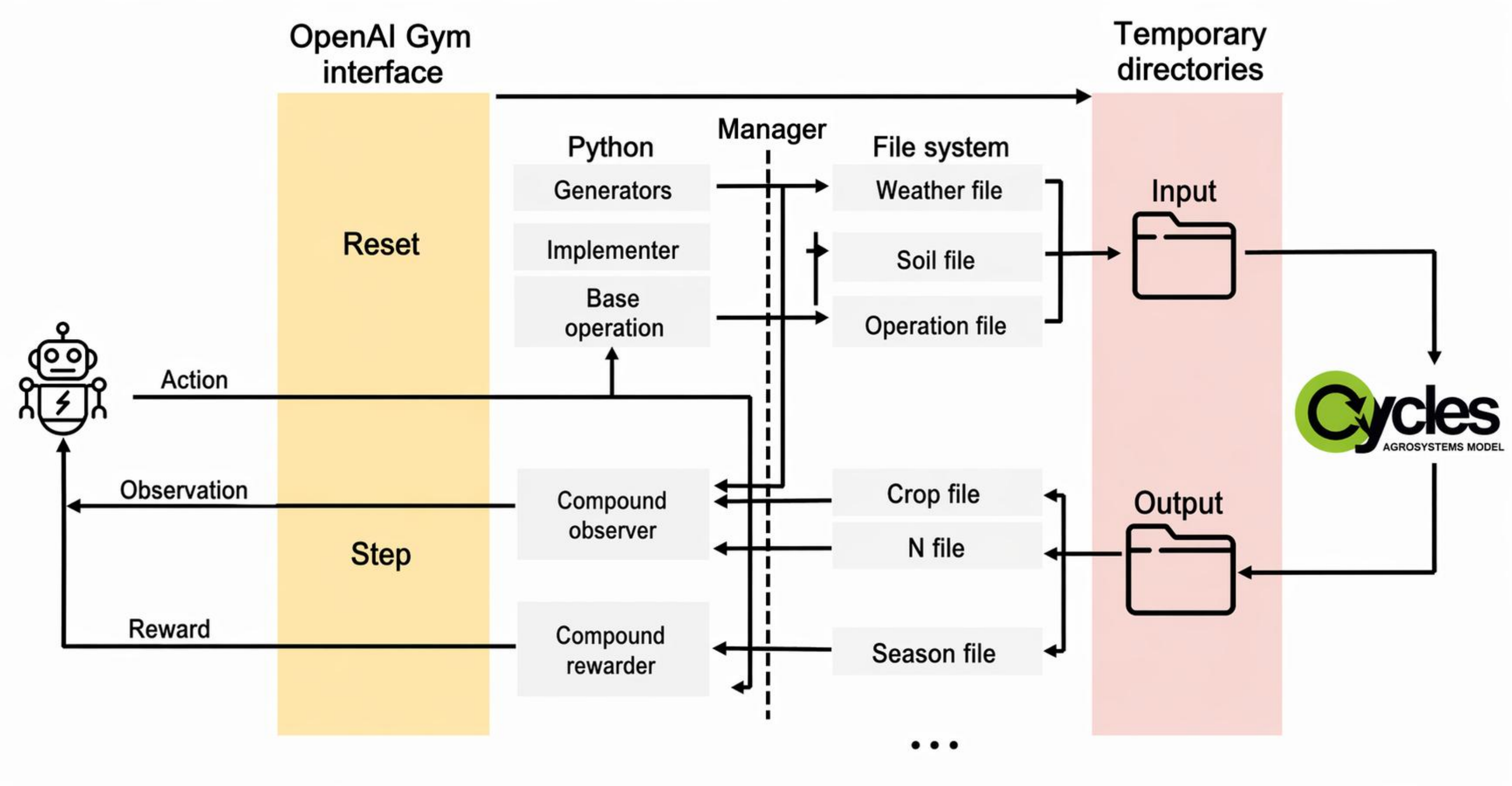
$$\mathcal{L}^{CLIP}_{\theta_k}(\theta) = \underset{\tau\sim\pi_k}{\mathrm{E}}\left[\sum_{t=0}^{T}\left[\min(r_t(\theta)\hat{A}^{\pi_k}_t, \text{clip}\,(r_t(\theta), 1-\epsilon, 1+\epsilon)\,\hat{A}^{\pi_k}_t)\right]\right]$$

end for

METHODOLOGY

Architecture

- Interactive diagrams:
  - Detailed View: complete thesis architecture and contribution map
  - Single-Run View: simplified PPO execution flow for fast interpretation



METHODOLOGY

One Step of RL Training - Agent and Environment (Steps 1-7)

1. Reset the environment - load Pakistan weather / soil / crop files; build first observation s_0 (fixed weather = controlled baseline, shuffled = stochastic training).

2. Build observation s_t - normalised vector of weather + crop (stage, biomass, N-stress, water-stress) + soil-N + time. Prices are used in the reward, not observed.

3. Policy forward pass - actor-critic MLP turns s_t into an action distribution pi(a|s) and a value estimate V(s).

4. Sample NPK action a_t - multi-discrete (N, P, K) with 11x11x11 = 1,331 combinations; N up to 150, P up to 80, K up to 60 kg/ha. Irrigation is fixed, not learned.

5. Step CyclesGym - write fertiliser operations; Cycles simulates the next 7 days (soil water, N-cycling, crop growth).

6. Build next state s_t+1 - parse new Cycles outputs into the same observation shape; episode is done at end of the season / year.

7. Compute reward r_t - compound signal: harvest revenue (CropRewarder) minus NPK fertiliser cost (NPKProfitabilityRewarder).

METHODOLOGY

One Step of RL Training - Collect Rollout and PPO Update (Steps 8-14)

8. Store transition - s_t, a_t, r_t, V(s_t), log pi_old(a_t|s_t) and the done flag go into the on-policy rollout buffer.

9. Collect a rollout - repeat steps 2-8 until the buffer is full (~2,048 steps, roughly 39 one-year episodes per env). Policy is frozen as theta_old.

10. GAE advantages - estimate A_t: did the action beat the critic's expectation? Blends short- and long-horizon TD errors (gamma~0.99, lambda~0.95).

11. PPO clipped objective - clip the policy ratio to within +/- epsilon (=0.2) so a single update cannot move the policy too far.

12. Full loss - clipped policy term + value-function loss + optional entropy bonus (ent_coef = 0 in the reported config).

13. Mini-batch SGD - shuffle the buffer; batch 64; 10 epochs; Adam lr 3e-4. Approximate KL is monitored as a stability diagnostic.

14. Clear and repeat - discard the buffer, set theta_old := theta_new. Budgets of 1k / 3k / 5k training-years ~ 53K / 159K / 265K env steps.

Experimentation

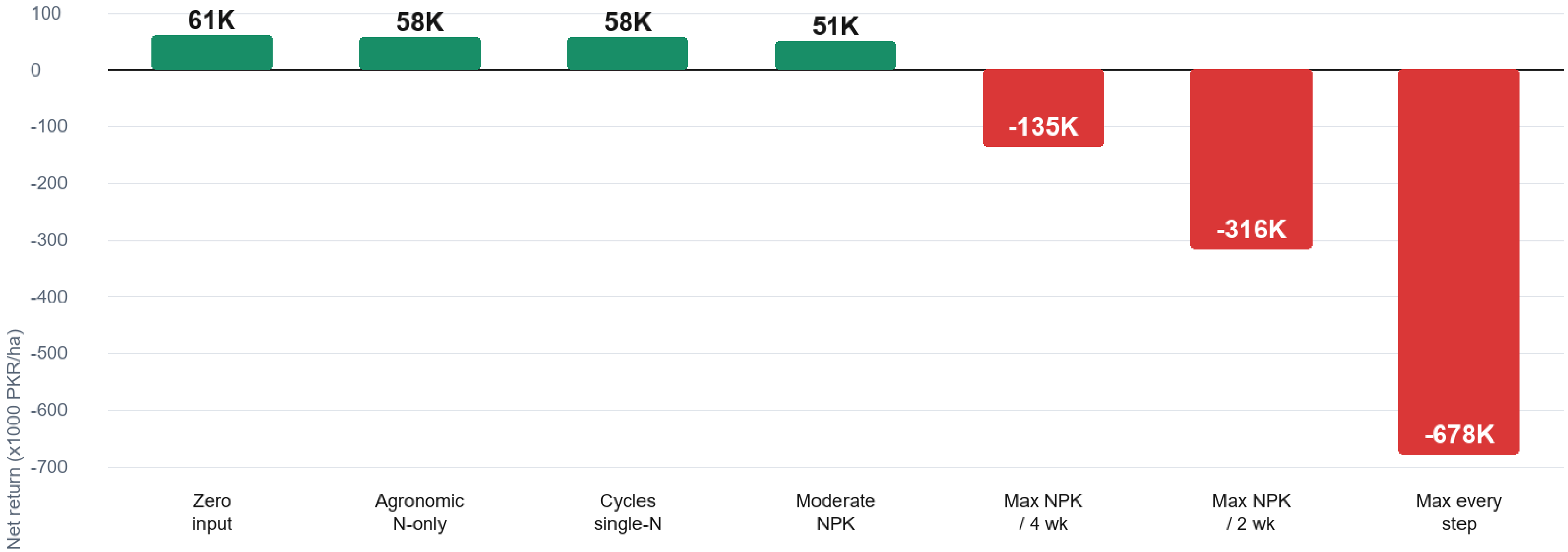
EXPERIMENTATION

Core Experiment Design

Algorithm	Objective Domain	Policy Mode	Weather Mode	Purpose in Thesis
PPO *(BEST RESULTS)*	Standalone Fertilization	Adaptive	Stochastic (Random)	To evaluate robust learning under real-world weather uncertainty.
PPO	Standalone Fertilization	Non-Adaptive	Fixed (Deterministic)	To establish a controlled baseline for the RL agent's learning capacity.
PPO	Hierarchical Crop Planning	Adaptive	Stochastic (Random)	To test multi-level decision making (crop choice + fertilization).
A2C	Standalone Fertilization	Adaptive	Stochastic (Random)	To benchmark PPO against an alternative Advantage-based method.
DQN	Standalone Fertilization	Adaptive	Fixed (Deterministic)	To justify the use of continuous/large action spaces over discrete ones.

EXPERIMENTATION

Fixed-Policy Baseline Stress Test



Baseline readout

- Fixed schedules are testable baselines, but not adaptive policies.
- Max-NPK schedules collapse cost-aware return.

Why this answers the examiner question

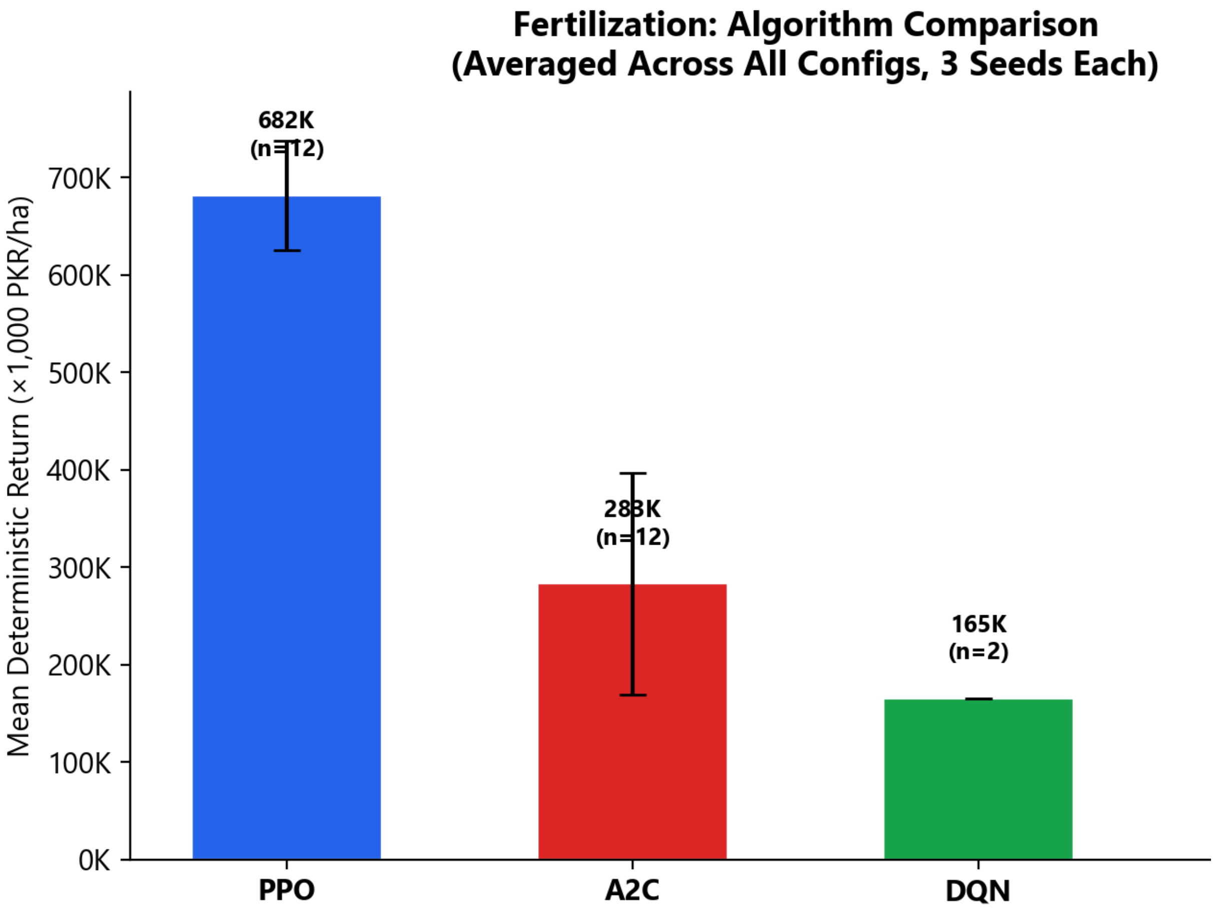
- The task needs sequential state-action feedback, not one fixed plan.
- RL learns timing/amount under simulator feedback.

EXPERIMENTATION

Results: Algorithm Comparison (Fertilization)

PPO gives the strongest single fertilization run in this simulation matrix

- Best PPO return: approx. 682K PKR/ha in the reported comparison
- A2C and DQN provide standard RL comparison/ablation baselines
- Interpret as simulated cost-aware return, not field yield improvement



EXPERIMENTATION

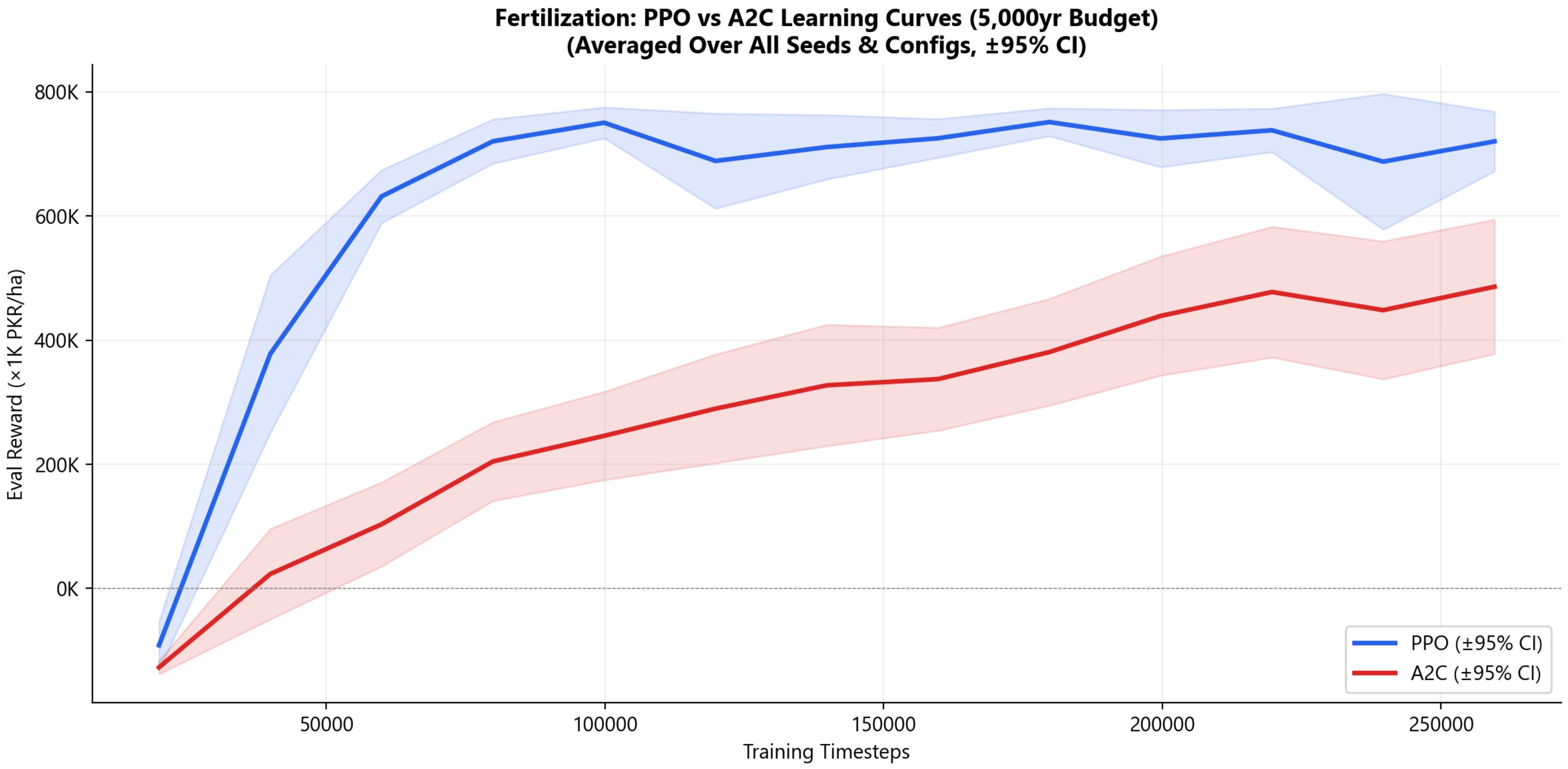
Results: Training Budget & Convergence

Sample Efficiency & Learning Curves

• PPO reaches massive convergence (plateaus) by 3,000 training years.

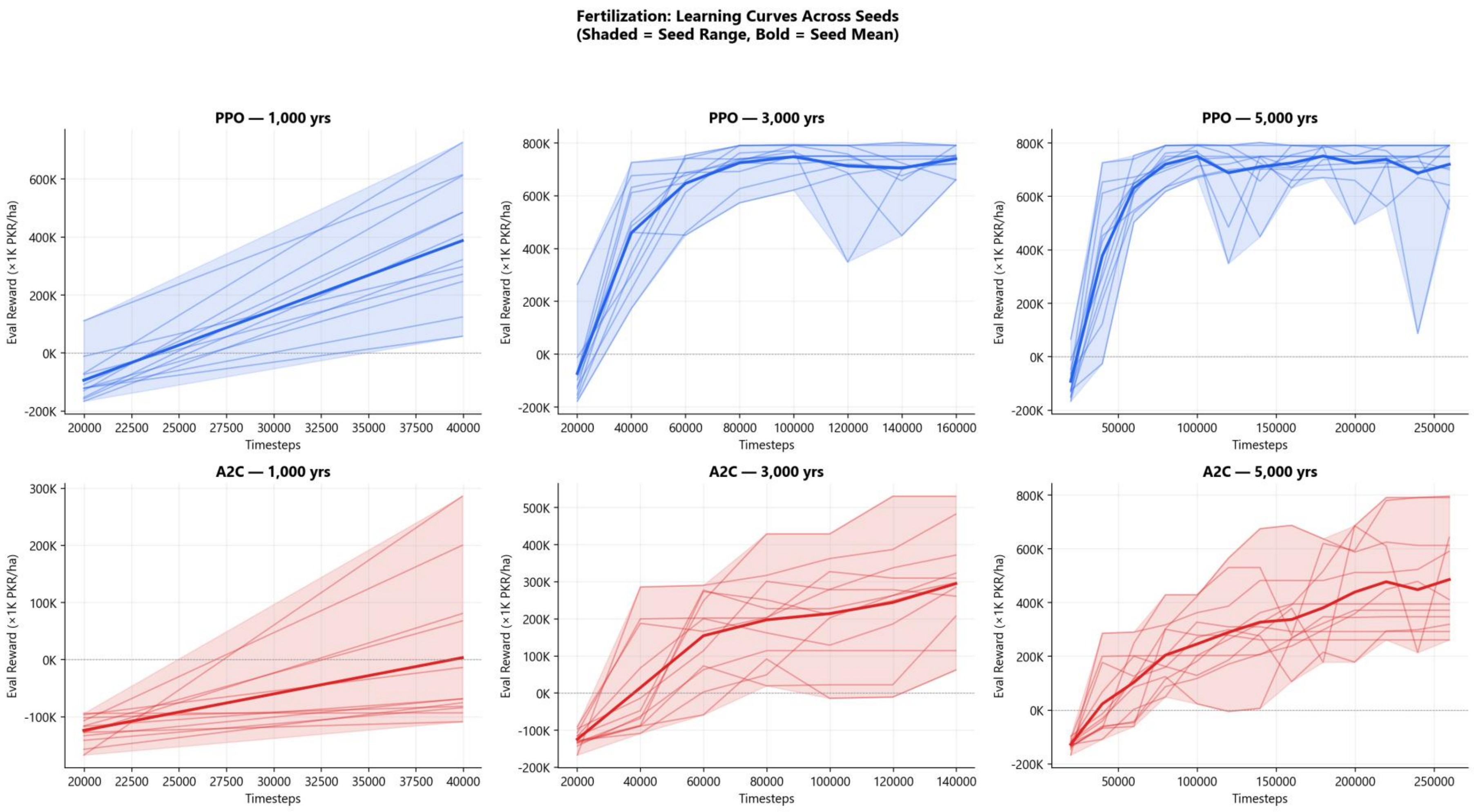
• A2C continues to struggle for efficiency, requiring far more data.

• Head-to-Head: PPO reaches 700K PKR by 80K timesteps; A2C reaches only 480K by 260K timesteps.



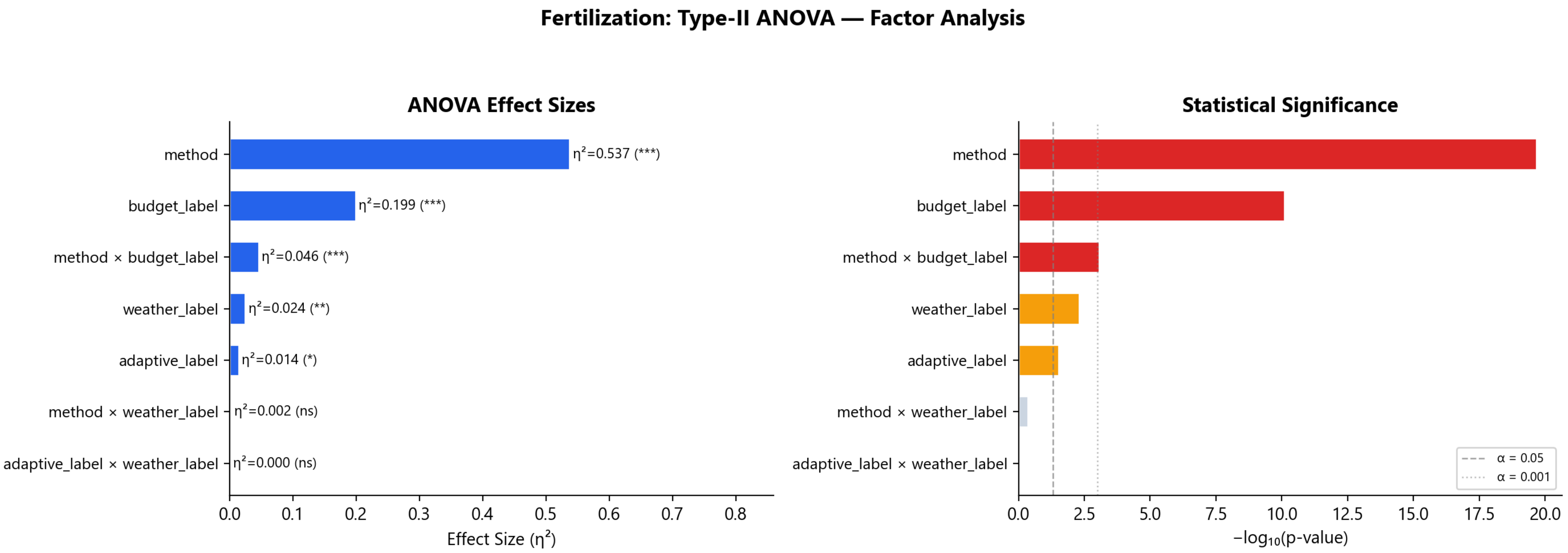
EXPERIMENTATION

Results: Training Budget & Convergence



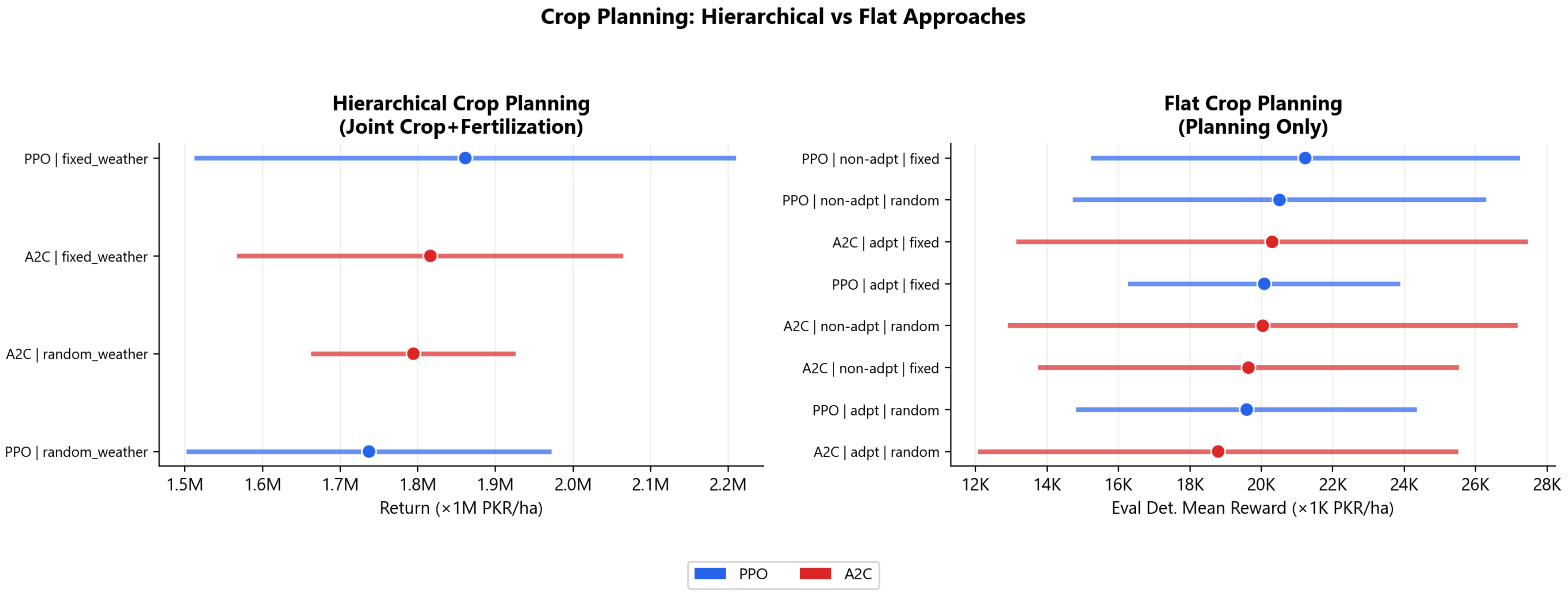
EXPERIMENTATION

Results: ANOVA Factor Significance



EXPERIMENTATION

Results: Hierarchical / Flat Scale Audit



Hierarchical and flat planning returns use different reward definitions, so raw magnitudes should not be read as direct superiority.

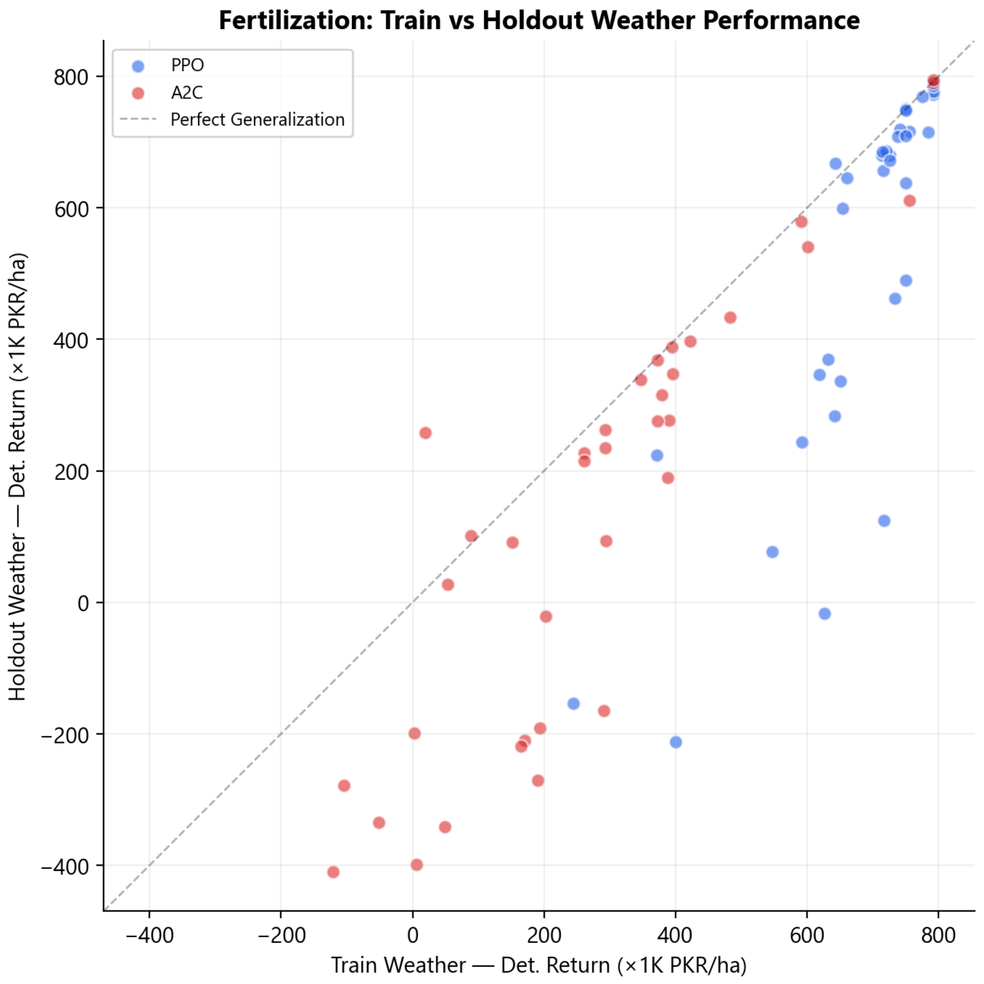
EXPERIMENTATION

Results: Generalization to Unseen Weather

Robustness Check: Holdout Weather Testing

• Testing the models on out-of-sample weather (2019+ Monsoons not seen in training).

• Result: PPO performance on Holdout weather mirrors its Training performance. No overfitting recorded.



LIMITATIONS AND FUTURE WORK

LIMITATIONS AND FUTURE WORK

LIMITATIONS

- Results are simulation-based only; no field trials or farmer-level validation were conducted.

- Active crop setup is maize-soy; rice-wheat, cotton, and multi-crop Pakistan systems are not tested.

- Evidence uses one audited Pakistan weather-soil site; multi-zone transfer remains untested.

- Irrigation is fixed, not a learned action; water-efficiency claims are out of scope.

- Seed count is limited (n = 3), so fine-grained stochastic comparisons should be treated cautiously.

- Hierarchical and flat returns use different reward scales; raw magnitudes are not directly comparable.

LIMITATIONS AND FUTURE WORK

FUTURE WORK

- Add **irrigation scheduling** as an RL control variable for water-use optimization.
- Extend the framework to model **biotic stress**, including pest, disease, and weed pressure.
- Add explicit **abiotic stress scenarios** such as drought, heatwaves, salinity, flooding, and waterlogging.
- Calibrate and validate the simulator using **local field data**, soil tests, yield records, and farmer management logs.
- Integrate other crop growth models such as **DSSAT, APSIM, AquaCrop, or ORYZA** for crop- and stress-specific studies.
- Develop multi-objective rewards covering **profit, yield stability, water use, nutrient losses, and emissions**.

CONCLUSION

CONCLUSION

Simulation-Based Decision-Support Potential

- Demonstrated simulation-based decision-support potential for cost-aware crop resource allocation.
- Localized CyclesGym with Pakistan weather, soil, crop-calendar, crop-price, and fertilizer-price inputs.
- Extended the nitrogen-focused workflow to NPK action/reward reporting and fixed-policy baseline stress tests.
- Evaluated standard PPO, A2C, and DQN; novelty is localization, reward design, hierarchy, reporting, and auditability.
- Maintains claim boundary: maize-soy setup, fixed irrigation, one audited site, n = 3 seeds, and no field-deployment claim.

Prototype Demonstration

THANK YOU

REFERENCES

[1] **Dhillon, M. S.**, T. Dahms, C. Kuebert-Flock, T. Rummler, J. Arnault, I. Steffan-Dewenter, and T. Ullmann. *“Integrating random forest and crop modeling improves the crop yield prediction of winter wheat and oil seed rape.”* Frontiers in Remote Sensing, vol. 3, 2023 (published Jan 2023).

[2] **Tabar, M.**, D. Lee, D. P. Hughes, and A. Yadav. *“Mitigating Low Agricultural Productivity of Smallholder Farms in Africa: Time-Series Forecasting for Environmental Stressors.”* Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI-22), 2022.

[3] **Uyeh, D. D.**, O. Iyiola, R. Mallipeddi, S. A. Hiablie, M. Amaizu, Y. Ha, and T. Park. *“Grid Search for Lowest Root Mean Squared Error in Predicting Optimal Sensor Location in Protected Cultivation Systems.”* Frontiers in Plant Science, vol. 13, 2022.

[4] **Shastri, S.**, S. Kumar, V. Mansotra, *et al. “Advancing crop recommendation system with supervised machine learning and explainable artificial intelligence.”* Scientific Reports, vol. 15, Article 25498, 2025.

[5] **Steininger, D.**, A. Trondl, G. Croonen, J. Simon, and V. Widhalm. *“The CropAndWeed Dataset: A Multi-Modal Learning Approach for Efficient Crop and Weed Manipulation.”* Proceedings of IEEE/CVF WACV, pp. 3729–3738, 2023.

REFERENCES

[6] **Cao, X.**, Y. Yao, L. Li, W. Zhang, Z. An, Z. Zhang, L. Xiao, *et al.* *"iGrow: A Smart Agriculture Solution to Autonomous Greenhouse Control."* Proceedings of AAAI 36(11), pp. 11837–11845, 2022.

[7] **Wu, J.**, D. Pichler, D. Marley, N. Hovakimyan, and D. A. Wilson. *"Extended Agriculture-Vision: An Extension of a Large Aerial Image Dataset for Agricultural Pattern Analysis."* Transactions on Machine Learning Research (TMLR), published Apr 2023.

[8] **Khan, A. I.**, S. M. K. Quadri, S. Banday, and J. L. Shah. *"Deep diagnosis: A real-time apple leaf disease detection system based on deep learning."* Computers and Electronics in Agriculture, vol. 199, Article 107093, 2022.

[9] **Tavera, A.**, E. Arnaudo, C. Masone, and B. Caputo. *"Augmentation Invariance and Adaptive Sampling in Semantic Segmentation of Agricultural Aerial Images."* Proceedings of CVPR 2022 Workshops (AgriVision), pp. 1656–1665, 2022.

[10] **Eleojo, J. B.** *"Utilizing Vision-Language Models for Detection of Leaf-Based Diseases in Tomatoes."* Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39 no. 28 (AI for Social Impact track), pp. 29567–29569, 2025.

[11] **Kallenberg, M. G. J.**, H. Overweg, R. van Bree, and I. N. Athanasiadis. *"Nitrogen management with reinforcement learning and crop growth models."* Environmental Data Science, vol. 2, e34, 2023.

REFERENCES

[12] **Gautron, R.**, O.-A. Maillard, P. Preux, M. Corbeels, and R. Sabbadin. *"Reinforcement Learning for Crop Management Support: A Review, Prospects and Challenges."* Computers and Electronics in Agriculture, vol. 200, Article 107182, 2022.

[13] **Turchetta, M.**, L. Corinzia, S. Sussex, A. Burton, J. Herrera, I. Athanasiadis, J. M. Buhmann, and A. Krause. *"Learning long-term crop management strategies with CyclesGym."* Advances in Neural Information Processing Systems 35 (NeurIPS Datasets and Benchmarks Track), 2022.

[14] **Avigal, Y.**, W. Wong, M. Presten, M. Theis, S. Aeron, A. Deza, S. Sharma, R. Parikh, S. Oehme, S. Carpin, J. H. Viers, S. Vougioukas, and K. Y. Goldberg. *"Simulating polyculture farming to learn automation policies for plant diversity and precision irrigation."* IEEE Transactions on Automation Science and Engineering, vol. 19 no. 3, pp. 1352–1364, 2022.

[15] **Din, A.**, M. Y. Ismail, B. Shah, M. Babar, F. Ali, and S. U. Baig. *"A deep reinforcement learning-based multi-agent area coverage control for smart agriculture."* Computers and Electrical Engineering, vol. 101, Article 108089, 2022.

[16] **Saikai, Y.**, A. Peake, and K. Chenu. *"Deep reinforcement learning for irrigation scheduling using high-dimensional sensor feedback."* PLOS Water, vol. 2 no. 9, e0000169, 2023.

REFERENCES

[17] **Wang, Y.**, Z. He, D. Cao, L. Ma, and Y. Cui. *“Coverage path planning for kiwifruit picking robots based on deep reinforcement learning.”* Computers and Electronics in Agriculture, vol. 205, Article 107593, 2023.

[18] **Vázquez-Canteli, J. R.**, U. R. Siddiqui, H. Madsen, and Z. Nagy. *“Reinforcement Learning versus Model Predictive Control on greenhouse climate control.”* Computers and Electronics in Agriculture, vol. 215, Article 108372, 2023.

[19] **Huang, S.**, A. Abdolmaleki, G. Vezzani, P. Brakel, D. J. Mankowitz, M. Neunert, S. Bohez, *et al. “A Constrained Multi-Objective Reinforcement Learning Framework.”* Proceedings of the 5th Conference on Robot Learning (CoRL), PMLR vol. 164, pp. 883–893, 2022.

[20] **Zhao, W.**, T. He, R. Chen, T. Wei, and C. Liu. *“State-wise Safe Reinforcement Learning: A Survey.”* Proceedings of the 32nd International Joint Conference on AI (IJCAI-23), Survey Track, pp. 6814–6820, 2023.