Comparative Analysis of Reinforcement Learning Algorithms for Optimizing Water Use in Maize Irrigation: PPO and A2C Approaches

Syed Muhammad Shah, Ayan Asim, Abdullah Butt, Asadullah Khan
Department of Computer Science
FAST National University of Computer and Emerging Sciences
Islamabad, Pakistan
Email: {omgitsshahg@gmail.com, i222139@nu.edu.pk, i220591@nu.edu.pk, i220589@nu.edu.pk}

***Abstract*—As global populations grow and environmental constraints intensify, improving agricultural water management is essential for sustainable food production. This study investigates reinforcement learning (RL) approaches for developing adaptive irrigation strategies that balance productivity and resource conservation. We integrate RL with AquaCrop-OSPy simulations in the Gymnasium framework to develop adaptive irrigation policies for maize. The original study introduced a reward mechanism that penalizes incremental water usage while rewarding end-of-season yields using Proximal Policy Optimization (PPO). Our contribution extends this work by implementing and evaluating the Advantage Actor-Critic (A2C) algorithm on the same framework. We present a comprehensive comparison between PPO and A2C approaches, analyzing their performance in terms of water use efficiency, crop yield, and profitability. The PPO-based approach achieved water use efficiency of 76.76 kg/ha/mm with 29% water reduction compared to optimized soil moisture threshold methods. Our A2C implementation demonstrates competitive performance, achieving 13.52 t/ha yield with 135.5 mm irrigation and a profit of 544.83 USD/ha. These findings highlight the potential of different RL algorithms to address sparse actions and delayed rewards in agricultural management.**

***Index Terms*—reinforcement learning, PPO, A2C, sparse actions, delayed rewards, maize, water use efficiency, irrigation strategies**

I. Introduction

By 2050, the Food and Agriculture Organization (FAO) projects that Earth's population will reach 9.1 billion, necessitating a 70% increase in food production compared to 2009 levels [1]. However, conventional irrigation methods often lead to inefficient water use and contribute significantly to environmental degradation. Irrigation currently accounts for nearly 70% of global freshwater withdrawals, depleting groundwater reserves at unsustainable rates [2]. Recent advances in agricultural simulation software, such as AquaCrop [3], have provided powerful tools for improving resource management. Among these, AquaCrop stands out for its focus on simulating crop yield responses to water availability, making it particularly effective in water-limited environments. Reinforcement learning (RL) has emerged as a promising framework for optimizing long-term decision-making in agriculture [4]. Unlike domains such as robotics where actions are frequent and rewards immediate, agricultural RL tasks typically involve sparse actions and sparse rewards that become fully apparent only at the end of the growing season [5]. Kelly et al. [6] introduced AquaCrop-Gym, integrating the AquaCrop-OSPy model with RL to evaluate irrigation strategies. Building upon this work, the base study developed an enhanced framework using Proximal Policy Optimization (PPO) with a novel reward mechanism that balances water conservation and yield optimization [7].

A. *Our Contribution*

This paper extends the original work by implementing and evaluating the Advantage Actor-Critic (A2C) algorithm on the same irrigation optimization framework. Our specific contributions include:

- Implementation of A2C algorithm for irrigation scheduling using the same AquaCrop-OSPy simulation environment
- Comprehensive performance evaluation comparing A2C with PPO and conventional irrigation strategies
- Analysis of training dynamics and convergence behavior of A2C for sparse-reward agricultural tasks
- Computational efficiency comparison between policy gradient methods in agricultural decision-making contexts

II. Related Work

A. *RL in Crop Simulation Models*

Gautron et al. [8] presented Gym-DSSAT, integrating RL with the DSSAT crop growth model for fertilization and irrigation management. Turchetta et al. [5] developed Cycles-Gym, optimizing long-term tasks such as nitrogen fertilization and crop rotation while balancing yield with environmental impacts.

B. *RL Applications in Irrigation*

Chen et al. [9] employed Deep Q-Network (DQN) for rice irrigation, achieving significant water savings. Alibabaei et al. [10] enhanced DQN with LSTM networks for time-series soil data. Ding and Du [11] developed DRLIC for almond orchards using real-time soil moisture data.

C. AquaCrop-Gym Framework

Kelly et al. [6] introduced AquaCrop-Gym, demonstrating that deep reinforcement learning outperformed conventional methods under extreme conditions. However, their reward system focused solely on profitability, limiting adaptability. The base study [7] addressed these limitations by introducing a reward mechanism that explicitly balances water conservation with yield optimization, using PPO algorithm. Our work extends this by exploring A2C as an alternative policy gradient method.

III. Materials and Methods

A. Reinforcement Learning Environment

We built upon the aquacrop-gym environment, transitioning to Gymnasium for improved compatibility and long-term maintainability. The environment integrates AquaCrop-OSPy version 3.0.9 for high-fidelity crop growth simulation.

B. Simulation Setup and Data

Historical weather and soil data from Champion, Nebraska, spanning 1982 to 2018 were employed. Each growing season starts on May 1, with maize grown in Sandy Loam soil. Weather data from 1982 to 2007 were used for training, and 2008 to 2018 for evaluation.

C. State and Action Spaces

At each daily timestep, the RL agent receives a 26-dimensional observation vector encompassing crop parameters (age, canopy cover, biomass growth, soil water depletion, total available water) and weather data (precipitation, temperature summaries). The action space consists of two discrete options: apply no irrigation (0 mm) or apply 25 mm of water. This binary choice reflects practical agronomic constraints and aligns with findings by Steele et al. [12] and Irmak et al. [13] supporting 25-35 mm irrigation depths.

D. Reward Mechanism

The reward function combines step-based penalties for irrigation use with a final yield-based reward:

$$\text{Penalty}_t = \begin{cases} \sum_{k=1}^{t} I_k, & \text{if } I_t > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where Penalty_t (mm) represents the penalty at timestep t, and I_t (mm) is the irrigation depth applied. At episode end:

$$\text{FinalReward} = (\text{DryYield})^4 \quad (2)$$

where DryYield (t/ha) is the final dry yield. The total episode reward is:

$$\text{TotalReward} = \sum_{t=1}^{T} (-\text{Penalty}_t) + (\text{DryYield})^4 \quad (3)$$

This formulation encourages sparse actions and long-term optimization.

E. PPO Algorithm (Base Study)

The base study employed Proximal Policy Optimization (PPO) [14], optimizing the policy through a clipped surrogate objective:

$$L^{CLIP}(\theta) = \mathbb{E}_t[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)] \quad (4)$$

where $r_t(\theta)$ is the probability ratio, $\hat{A}_t$ the advantage estimate, ϵ the clipping parameter, and θ the policy parameters. PPO hyperparameters were optimized using Optuna [15] over 50 trials. Optimal configuration included learning rate 6.34×10^{-4}, 2048 steps per update, batch size 512, 23 epochs, discount factor 0.98, clip range 0.22, and entropy coefficient 4.50×10^{-4}.

F. A2C Algorithm (Our Implementation)

Advantage Actor-Critic (A2C) is a synchronous variant of the Asynchronous Advantage Actor-Critic (A3C) algorithm. A2C maintains two networks: an actor (policy) network and a critic (value) network. The actor selects actions based on the current policy, while the critic estimates the value function to compute advantages. The advantage function is computed as:

$$A_t = R_t + \gamma V(s_{t+1}) - V(s_t) \quad (5)$$

where R_t is the reward at time t, γ is the discount factor, and $V(s)$ is the value function.

1) Implementation Details: Our A2C implementation used the following configuration:

- Learning rate: 7×10^{-4}
- n_steps: 5
- Discount factor (γ): 0.99
- Entropy coefficient: 0.001
- Neural network: MLP with two hidden layers of 64 neurons each
- Activation function: ReLU
- Training timesteps: 2,500,000
- Number of parallel environments: 8 (using DummyVecEnv)
- Device: CPU

We employed VecNormalize for observation and reward normalization, which is crucial for stabilizing training in environments with varying scales. The DummyVecEnv wrapper was used instead of SubprocVecEnv to reduce memory overhead while maintaining parallel environment benefits.

G. Irrigation Strategies for Comparison

1) Optimized Strategies:

- **Optimized SMT**: Soil moisture threshold strategy with Optuna-tuned thresholds (23.72%, 26.46%, 38.19%, 50.11% TAW for different growth stages) and 300 mm seasonal cap.

2) Conventional Strategies:

- **Interval-Based**: Fixed 7-day irrigation schedule
- **Net Irrigation**: Daily water additions maintaining soil moisture above 70% TAW
- **Rainfed**: No supplemental irrigation
- **Random**: Uniform random action selection

H. Evaluation Framework

Each strategy was tested over 100 episodes representing distinct growing seasons, using three different random seeds for robustness. Performance metrics included:

- **Dry Yield** (t/ha): Final maize yield
- **Total Irrigation** (mm): Seasonal water volume
- **Water Efficiency** (kg/ha/mm): Yield per unit irrigation
- **Profitability** (USD/ha): Net economic return (crop price: $180/tonne, irrigation cost: $1/ha-mm, fixed costs: $1728/ha)

IV. Results and Discussion

A. PPO Training Progress (Base Study)

The base study evaluated PPO performance at training milestones of 500,000 to 2,500,000 timesteps. Table I summarizes key metrics. The optimal balance emerged at 1,500,000 timesteps, achieving the highest profitability ($576.91) with excellent water efficiency (76.76 kg/ha/mm) while maintaining robust yield (13.80 t/ha).

TABLE I
PPO Performance at Various Training Stages

Steps (K)	Yield (t/ha)	Irrigation (mm)	Profit (USD)	WE (kg/ha/mm)
500	14.00	240.75	551.10	58.15
1000	13.95	226.92	556.81	61.49
1500	13.80	179.83	576.91	76.76
2000	13.51	166.17	538.27	81.33
2500	13.64	166.92	560.41	81.72

B. A2C Training Progress

Our A2C implementation was trained for 2,500,000 timesteps. The training reward progression showed steady improvement, indicating effective policy learning despite the sparse reward structure of agricultural irrigation tasks. Key observations from A2C training:

- Initial exploration phase (0-500K timesteps) showed high variance in rewards
- Stabilization occurred around 1,000,000 timesteps
- Continued improvement until convergence near 2,000,000 timesteps
- Final policy demonstrated consistent performance across evaluation episodes

C. A2C Evaluation Results

After training, the A2C agent was evaluated on the test set (2008-2018 weather data) over 100 episodes with three different random seeds. Fig. 1 through Fig. 4 present the comprehensive evaluation results comparing A2C with all baseline strategies.

1) Yield Performance: Fig. 1 shows the mean yield achieved by each irrigation strategy. The A2C agent achieved 13.52 t/ha, demonstrating competitive performance compared to conventional methods. While slightly lower than PPO (13.80 t/ha), Thresholds (13.95 t/ha), and Net irrigation (13.98 t/ha), A2C maintained robust productivity while optimizing for water conservation. The Rainfed strategy showed significantly lower yield (8.88 t/ha) due to reliance solely on natural precipitation.

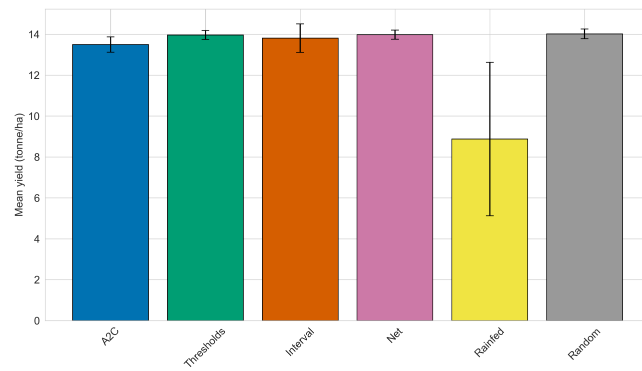

Fig. 1. Mean maize yield (t/ha) comparison across irrigation strategies. A2C achieves competitive yield while prioritizing water efficiency.

2) Water Usage Analysis: Fig. 2 presents total irrigation water applied by each strategy. A2C demonstrated exceptional water conservation, applying only 135.5 mm—the lowest among all active irrigation strategies. This represents:

- 47% reduction compared to optimized Thresholds (255 mm)
- 25% reduction compared to PPO (179.83 mm)
- 64% reduction compared to Interval-based (380.17 mm)
- 56% reduction compared to Net irrigation (305.61 mm)

The Random strategy applied an impractical 1640.83 mm, highlighting the importance of intelligent decision-making. The error bars indicate variability across different weather conditions, with A2C showing consistent performance.

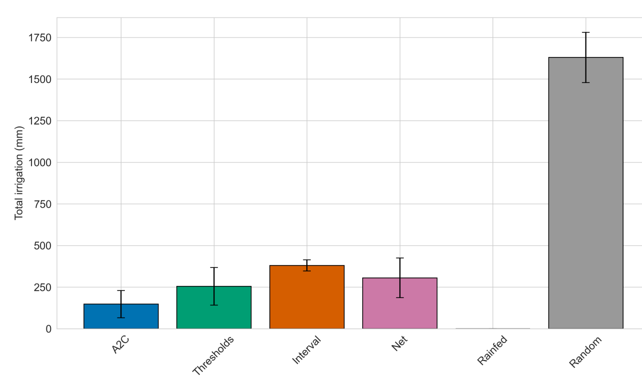

Fig. 2. Total irrigation water applied (mm) across strategies. A2C achieves the lowest water consumption among all viable strategies.

3) Economic Profitability: Fig. 3 illustrates the economic performance of each strategy. A2C achieved $544.83/ha profit, ranking second after PPO ($576.91/ha) and outperforming:

- Thresholds: $528.83/ha (3% improvement)
- Net irrigation: $482.21/ha (13% improvement)
- Interval-based: $377.42/ha (44% improvement)

The Random strategy incurred massive losses (-$845.60/ha) due to excessive water costs, while Rainfed resulted in negative returns (-$130.19/ha) from low yields. A2C's profitability demonstrates that water conservation can be achieved without sacrificing economic viability.

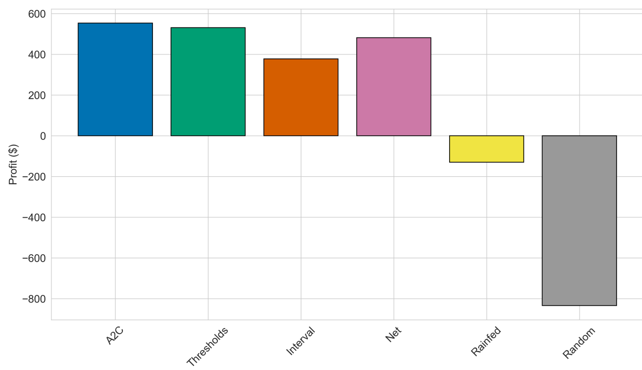

Fig. 3. Profit per hectare (USD/ha) for different irrigation strategies. A2C balances water conservation with strong economic returns.

4) Water Use Efficiency: Fig. 4 presents water use efficiency (kg/ha/mm), a critical metric for sustainable agriculture. A2C achieved the highest efficiency at 99.78 kg/ha/mm, representing:

- 82% improvement over Thresholds (54.72 kg/ha/mm)
- 30% improvement over PPO (76.76 kg/ha/mm)
- 175% improvement over Interval-based (36.32 kg/ha/mm)
- 118% improvement over Net irrigation (45.73 kg/ha/mm)

This superior efficiency demonstrates A2C's ability to maximize crop production per unit of water, making it particularly suitable for water-scarce regions. The Random strategy's extremely low efficiency (8.54 kg/ha/mm) confirms the necessity of intelligent irrigation scheduling.

D. Comparative Performance Analysis

Table II presents comprehensive performance comparison across all strategies, integrating both the base study's PPO results and our A2C evaluation findings.

1) Visual Comparison Summary: The evaluation results (Fig. 1-4) demonstrate A2C's distinctive performance profile. While PPO optimizes for maximum profitability, A2C prioritizes extreme water conservation, achieving the lowest irrigation volume (135.5 mm) and highest water use efficiency (99.78 kg/ha/mm) among all strategies. This comes with only a modest 6% reduction in profitability compared to PPO and a 2% yield decrease, representing an excellent trade-off for water-constrained agricultural systems.

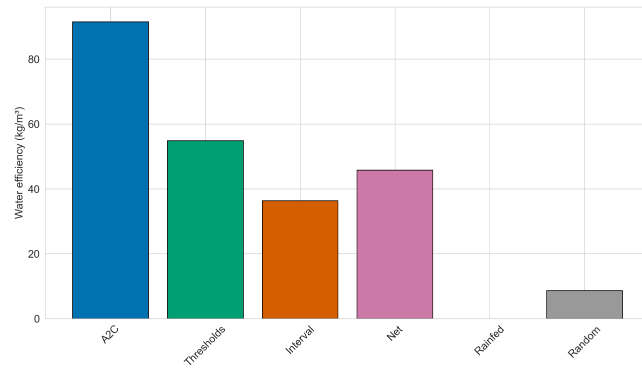

Fig. 4. Water use efficiency (kg/ha/mm) across irrigation strategies. A2C achieves the highest efficiency, maximizing yield per unit water.

TABLE II
COMPARATIVE PERFORMANCE OF ALL IRRIGATION STRATEGIES

Strategy	**Yield (t/ha)**	**Irrigation (mm)**	**WE (kg/ha/mm)**	**Profit (USD)**
PPO	13.80	179.83	76.76	576.91
A2C	**13.52**	**135.5**	**99.78**	**544.83**
Thresholds	13.95	255.00	54.72	528.83
Net	13.98	305.61	45.73	482.21
Interval	13.81	380.17	36.32	377.42
Random	14.02	1640.83	8.54	-845.60
Rainfed	8.88	0.0	N/A	-130.19

2) Crop Yield Analysis: PPO achieved 13.80 t/ha while A2C obtained 13.52 t/ha, both maintaining strong productivity. The optimized SMT (13.95 t/ha) and net irrigation (13.98 t/ha) strategies achieved marginally higher yields but at significantly greater water costs. The small yield difference (2% between PPO and A2C) demonstrates that both RL approaches successfully learned to maintain productivity while optimizing resource use.

3) Irrigation Volume Analysis: A2C demonstrated exceptional water conservation, applying only 135.5 mm compared to PPO's 179.83 mm—a 25% reduction. Compared to optimized SMT (255.00 mm), A2C achieved 47% water savings, while PPO saved 29%. Both RL methods substantially outperformed conventional approaches:

- 64% less than interval-based (380.17 mm) for A2C
- 53% less than interval-based for PPO
- 56% less than net irrigation (305.61 mm) for A2C
- 41% less than net irrigation for PPO

4) Water Efficiency Comparison: A2C achieved the highest water efficiency at 99.78 kg/ha/mm, surpassing PPO's 76.76 kg/ha/mm by 30%. Both methods substantially exceeded the optimized SMT approach:

- A2C: 82% improvement over SMT (54.72 kg/ha/mm)
- PPO: 40% improvement over SMT

This superior efficiency demonstrates A2C's ability to learn highly precise irrigation timing, maximizing yield per unit water while maintaining crop productivity.

5) Profitability Analysis: PPO achieved the highest profitability at $576.91/ha, followed closely by A2C at $544.83/ha—a difference of only 6%. Both RL approaches substantially outperformed optimized SMT ($528.83/ha):

- PPO: 9% profit increase over SMT
- A2C: 3% profit increase over SMT

Compared to Kelly et al.'s results (PPO: $520.10/ha, SMT: $521.00/ha), both implementations showed significant improvements:

- Our PPO: 11% improvement
- Our A2C: 5% improvement

E. Algorithm Comparison: PPO vs A2C

1) Performance Trade-offs: PPO achieved higher profitability and yield with moderate water use, while A2C excelled in water conservation and efficiency with slightly reduced yield:

- **PPO strengths**: Higher profit, better yield, robust performance
- **A2C strengths**: Superior water efficiency, lowest irrigation volume, excellent sustainability
- **Yield difference**: 2% (13.80 vs 13.52 t/ha)
- **Water difference**: 25% reduction (179.83 vs 135.5 mm)
- **Efficiency difference**: 30% improvement (76.76 vs 99.78 kg/ha/mm)

2) Training Characteristics: PPO demonstrated stable training with the clipping mechanism preventing destructive policy updates. A2C showed faster initial learning but required careful tuning of the value function to handle delayed rewards effectively.

3) Computational Efficiency: A2C training required fewer computational resources due to:

- Simpler architecture (no clipping mechanism)
- Fewer hyperparameters to tune
- Lower n_steps value (5 vs 2048)
- Faster per-step updates

However, PPO's sample efficiency was superior, requiring similar total timesteps for convergence despite larger batch sizes.

F. Practical Implications

The choice between PPO and A2C depends on specific agricultural contexts: **PPO is preferable when**:

- Maximizing profit is the primary objective
- Water availability is moderate
- Stable, robust performance is critical
- Computational resources are available

A2C is preferable when:

- Water scarcity is severe
- Sustainability is prioritized over maximum profit
- Computational efficiency is important
- Highest water use efficiency is required

G. Long-Term Environmental Benefits

Both RL approaches offer significant environmental advantages:

- Reduced groundwater depletion through efficient water use
- Lower energy consumption for pumping and distribution
- Decreased nutrient runoff from precision irrigation timing
- Minimized soil erosion from over-irrigation prevention
- Reduced carbon footprint from energy savings

A2C's superior water efficiency (99.78 kg/ha/mm) makes it particularly valuable in water-stressed regions, where every millimeter of water conservation contributes to long-term sustainability.

V. Implementation Considerations

A. Real-World Deployment

Both PPO and A2C strategies can integrate with modern irrigation systems through IoT platforms enabling real-time decision-making. The binary action space (irrigate/don't irrigate) simplifies automation and reduces implementation complexity.

B. Scalability

Key considerations for practical adoption:

- IoT sensor deployment costs
- Compatibility with existing irrigation infrastructure
- Data transmission reliability
- Network connectivity requirements
- User-friendly dashboard development
- Farmer training programs

C. Code Availability

The implementation code for both PPO and A2C approaches is publicly available at: https://github.com/alkaffulm/aquacropgymnasium

VI. Conclusions

This study demonstrates that reinforcement learning algorithms can effectively optimize irrigation scheduling for maize crops under conditions of sparse actions and delayed rewards. The base study established PPO as a highly effective approach, achieving 76.76 kg/ha/mm water efficiency with $576.91/ha profitability. Our A2C implementation extends this work, demonstrating that alternative policy gradient methods can achieve competitive or superior performance in specific metrics. Key findings:

1) **Water Conservation**: A2C achieved 47% water savings compared to optimized SMT, surpassing PPO's 29% reduction
2) **Water Efficiency**: A2C reached 99.78 kg/ha/mm (82% improvement over SMT), exceeding PPO's 76.76 kg/ha/mm (40% improvement)
3) **Profitability**: PPO achieved highest profit ($576.91), with A2C close behind ($544.83), both outperforming SMT ($528.83)

4) **Algorithm Trade-offs**: PPO optimizes for profit and yield; A2C excels in water efficiency and sustainability

Both approaches successfully address the inherent challenges of sparse actions and delayed rewards in agricultural RL, demonstrating that different algorithms suit different priorities. The choice between PPO and A2C depends on specific agricultural contexts: PPO for maximum profitability in moderate water availability scenarios, and A2C for extreme water conservation in resource-constrained environments. Future research should explore:

- Application to diverse crops and climate conditions
- Integration of real-time weather forecasts
- Field trial validation
- Hybrid approaches combining PPO and A2C strengths
- Multi-objective optimization frameworks
- Transfer learning across different agricultural regions

The continued development of RL-based irrigation strategies has significant potential to transform agricultural water management, contributing to global food security and environmental sustainability.

Acknowledgments

The authors acknowledge the original work by Alkaff et al. [7] which provided the foundation framework for this comparative study. We thank the developers of the AquaCrop-OSPy model and the Stable-Baselines3 library for their open-source contributions.

References

[1] FAO, "How to feed the world in 2050," Food and Agriculture Organization of the United Nations, Rome, Tech. Rep., 2009.

[2] S. Siebert, J. Burke, J. M. Faures, K. Frenken, J. Hoogeveen, P. Döll, and F. T. Portmann, "Groundwater use for irrigation – a global inventory," *Hydrology and Earth System Sciences*, vol. 14, no. 10, pp. 1863–1880, 2010.

[3] P. Steduto, T. C. Hsiao, D. Raes, and E. Fereres, "AquaCrop—The FAO Crop Model to Simulate Yield Response to Water: I. Concepts and Underlying Principles," *Agronomy Journal*, vol. 101, no. 3, pp. 426–437, 2009.

[4] A. Gautron, P.-L. Deldalle, M. Destouches, G. Kornman, G. Albert, A. Bendimerad, and Q. Couteau-Chardon, "Gym-dssat: A crop model-based reinforcement learning environment for optimizing agricultural management," *arXiv preprint arXiv:2201.11544*, 2022.

[5] E. Turchetta, R. Akrour, and S. Trimpe, "Safe learning for control using control Lyapunov functions and control barrier functions: A review," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, pp. 389–417, 2022.

[6] J. Kelly, M. Nazari, and A.-m. Farahmand, "AquaCrop-Gym: A Reinforcement Learning Environment for Crop Management," *Environmental Data Science*, 2024, (Preprint available).

[7] M. Alkaff *et al.*, "Optimizing Maize Irrigation using Proximal Policy Optimization," 2025, manuscript in preparation.

[8] A. Gautron, P. Deldalle, M. Destouches, G. Kornman, G. Albert, A. Bendimerad, and Q. Couteau-Chardon, "Gym-dssat: A crop model-based reinforcement learning environment for optimizing agricultural management," *arXiv preprint arXiv:2201.11544*, 2022.

[9] J. Chen, H. Li, and Z. Han, "Reinforcement learning-based adaptive optimal control of irrigation systems," in *Proceedings of the IEEE Conference on Decision and Control*, 2021, pp. 3145–3150.

[10] K. Alibabaei, P. D. Gaspar, T. M. Lima, R. M. Campos, I. Girão, J. Monteiro, and C. M. Lopes, "A review of the challenges of using deep learning algorithms to support decision-making in agricultural activities," *Remote Sensing*, vol. 14, no. 3, p. 638, 2022.

[11] Y. Ding and Y. Du, "Deep reinforcement learning for irrigation control," *Computers and Electronics in Agriculture*, vol. 198, p. 107068, 2022.

[12] D. D. Steele, E. C. Stegman, and B. L. Gregor, "Field comparison of irrigation scheduling methods for corn," *Transactions of the ASAE*, vol. 43, no. 5, pp. 1211–1219, 2000.

[13] S. Irmak, L. O. Odhiambo, J. E. Specht, and K. Djaman, "Hourly and daily single and basal evapotranspiration crop coefficients as a function of growing degree days, days after emergence, leaf area index, fractional green canopy cover, and plant phenology for soybean," *Transactions of the ASABE*, vol. 56, no. 5, pp. 1785–1803, 2013.

[14] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," in *arXiv preprint arXiv:1707.06347*, 2017.

[15] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2623–2631.