Article

Optimizing Water Use in Maize Irrigation with Reinforcement Learning

Muhammad Alkaff [1,2], Abdullah Basuhail [1] and Yuslena Sari [2,*]

[1] Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; malkaff0001@stu.kau.edu.sa or m.alkaff@ulm.ac.id (M.A.); abasuhail@kau.edu.sa (A.B.)

[2] Department of Information Technology, Universitas Lambung Mangkurat, Banjarmasin 70123, Indonesia

* Correspondence: yuzlena@ulm.ac.id

Abstract: As global populations grow and environmental constraints intensify, improving agricultural water management is essential for sustainable food production. Traditional irrigation methods often lack adaptability, leading to inefficient water use. Reinforcement learning (RL) offers a promising solution for developing dynamic irrigation strategies that balance productivity and resource conservation. However, agricultural RL tasks are characterized by sparse actions—irrigation only when necessary—and delayed rewards realized at the end of the growing season. This study integrates RL with AquaCrop-OSPy simulations in the Gymnasium framework to develop adaptive irrigation policies for maize. We introduce a reward mechanism that penalizes incremental water usage while rewarding end-of-season yields, encouraging resource-efficient decisions. Using the Proximal Policy Optimization (PPO) algorithm, our RL-driven approach outperforms fixed-threshold irrigation strategies, reducing water use by 29% and increasing profitability by 9%. It achieves a water use efficiency of 76.76 kg/ha/mm, a 40% improvement over optimized soil moisture threshold methods. These findings highlight RL's potential to address the challenges of sparse actions and delayed rewards in agricultural management, delivering significant environmental and economic benefits.

Keywords: reinforcement learning; sparse actions; delayed rewards; proximal policy optimization; maize; water use efficiency; irrigation strategies

MSC: 68T20

1. Introduction

As global populations continue to grow and environmental challenges become more pronounced, optimizing water management in agriculture has emerged as a critical priority for sustainable food production. By 2050, the Food and Agriculture Organization (FAO) projects that the Earth's population will reach 9.6 billion, necessitating a 50% increase in food production compared to 2013 levels [1]. However, conventional irrigation methods, which prioritize maximizing crop yield, often lead to inefficient water use and contribute significantly to environmental degradation [2]. Irrigation currently accounts for nearly 70% of global freshwater withdrawals, depleting groundwater reserves at unsustainable rates [3]. This excessive water consumption threatens the sustainability of vital water resources and poses substantial risks to long-term food security [4,5]. Furthermore, the inability of traditional irrigation systems to adapt dynamically to changing weather patterns and fluctuating water availability exacerbates water wastage and highlights the inefficiency of

these approaches [6,7]. Addressing these challenges requires innovative irrigation strategies that prioritize resource conservation without compromising agricultural productivity [8,9]. Optimizing water use in agriculture is essential not only for environmental sustainability but also for meeting growing food demands in the face of increasing water scarcity [9].

Recent advances in agricultural simulation software have provided powerful tools for improving resource management and enabling more efficient farming practices. Examples of these tools include APSIM (Agricultural Production Systems Simulator) [10], DSSAT (Decision Support System for Agrotechnology Transfer) [11], and AquaCrop [12]. These simulations enable the accurate modeling of crop growth, environmental interactions, and resource management strategies, allowing farmers to test and refine practices in virtual environments before implementation in the field [13]. Among these tools, AquaCrop stands out for its focus on simulating crop yield responses to water availability, making it particularly effective in water-limited environments. By modeling the interactions between crop growth and water supply, AquaCrop offers valuable insights into how irrigation strategies impact yield under varying conditions.

In addition to agricultural simulation tools, artificial intelligence (AI) has emerged as a transformative force in agriculture, driving innovations in resource optimization and decision-making. Techniques such as deep learning (DL) and neural networks (NNs) are being applied to monitor plant health, predict water stress, and automate tasks like weeding and harvesting, improving operational efficiency [14]. AI-based automation, leveraging tools like artificial neural networks (ANNs) and the Internet of Things (IoT), enables the precise real-time monitoring of soil moisture and temperature, enhancing resource management and sustainability.

Building on these AI-driven advancements, reinforcement learning (RL) has emerged as a promising framework for optimizing long-term decision-making in agriculture [15]. In the context of agricultural water management, RL has the potential to address complex challenges by dynamically adapting to real-time environmental conditions. RL algorithms could optimize irrigation schedules by accounting for fluctuations in weather patterns, variations in soil moisture levels, and unforeseen rainfall events. However, unlike domains such as robotics, where actions are frequently taken and rewards are often immediate, agricultural RL tasks typically involve sparse actions—agents should only intervene (irrigate) when truly necessary—and sparse rewards that become fully apparent only at the end of the growing season [16]. This inherent sparsity in both actions and rewards introduces unique computational and learning challenges, requiring RL approaches tailored to the agricultural context. This adaptability offers the potential to enhance water use efficiency, minimize resource wastage, and support improved crop productivity under diverse and dynamic conditions.

The integration of RL with agricultural simulations has garnered considerable interest, with platforms such as CropGym [17,18], gym-DSSAT [19], CyclesGym [16], and SWAT-Gym [20] enabling RL models to make informed decisions based on simulated agricultural outcomes. For instance, gym-DSSAT [19] integrates RL with the DSSAT crop growth model [11] to simulate management tasks such as fertilization and irrigation, achieving higher yields with reduced resource input. Similarly, CyclesGym [16] integrates the Cycles crop model [21] to optimize long-term tasks such as nitrogen fertilization and crop rotation, balancing yield with environmental impacts such as nitrate leaching.

Expanding upon these prior developments, Kelly et al. (2024) introduced AquaCrop-Gym, a pioneering framework that integrates the AquaCrop-OSPy model [22] with reinforcement learning (RL) to evaluate irrigation strategies and their impact on crop yield. Their study demonstrated the potential of RL-driven irrigation, particularly under extreme conditions, where it outperformed conventional methods. Unlike Gym-DSSAT and Cycles-

Gym, which integrate RL for broader crop management tasks, AquaCrop-Gym specifically focuses on irrigation optimization using AquaCrop-OSPy's detailed water–crop interactions. However, its reward mechanism, centered solely on profitability, limits adaptability and real-world applicability. Moreover, the study highlighted the need for RL approaches that can better handle high rainfall variability, where conventional methods often proved more effective. These limitations provide an opportunity for further research to enhance the adaptability and efficiency of RL-based irrigation strategies.

Most RL-based studies, including AquaCrop-Gym, have focused primarily on yield or profit maximization. For instance, CropGym and CyclesGym integrate RL to optimize nutrient and water management but do not explicitly isolate irrigation as the primary optimization target. This approach can obscure the specific impact of irrigation optimization. Further research, such as [23–25], has applied RL to specific crops, demonstrating water savings without compromising yield, yet they lack a comprehensive focus on irrigation optimization.

Our study builds upon the framework established by Kelly et al. [26], addressing key gaps in adaptability, reward design, and evaluation. By isolating irrigation as the primary variable, we introduce a novel reward system that not only penalizes excessive water use through step penalties and incentivizes end-of-season yield optimization, but also inherently encourages sparse actions—irrigation interventions occur only when truly beneficial—aligning with the delayed and intermittent nature of agricultural decision-making. Unlike prior works, such as Gym-DSSAT and CyclesGym, which integrate multiple resource management tasks, our approach focuses exclusively on irrigation management, ensuring the precise evaluation of water use efficiency. This approach explicitly balances water conservation with agricultural productivity, overcoming the narrow profit-maximization focus of prior studies. Additionally, this study adopts Gymnasium [27], a stable and actively maintained RL environment, which builds on and improves OpenAI Gym [28], the framework used in Kelly et al. [26] AquaCrop-Gym. Gymnasium ensures long-term support, better compatibility with modern RL tools, and an improved user experience for developing and evaluating RL-based irrigation strategies. By integrating AquaCrop-OSPy simulations with Gymnasium, we provide a robust and flexible platform for testing and comparing irrigation strategies through reinforcement learning (RL).

We developed an adaptive irrigation management system for maize crops using Proximal Policy Optimization (PPO) [29], a state-of-the-art reinforcement learning algorithm widely recognized for its ability to handle continuous control tasks efficiently. PPO optimizes policies by balancing exploration and exploitation, ensuring stable and reliable learning in dynamic environments such as crop irrigation. The algorithm was trained within the Gymnasium framework, seamlessly interfacing with AquaCrop-OSPy to simulate real-world agricultural conditions. Building on the RL-based advancements seen in CropGym, gym-DSSAT, and AquaCrop-Gym, our system focuses explicitly on addressing real-world irrigation challenges while enhancing water use efficiency. It dynamically optimizes water use while maintaining high crop yields, achieving a critical balance between resource conservation and agricultural productivity. Furthermore, our approach is designed to address the unique challenges of RL in agriculture, where irrigation decisions are inherently sparse and rewards are delayed until the end of the growing season. By leveraging a tailored reward mechanism that penalizes excessive water use and rewards end-of-season yield, we ensure that the RL agent learns effective policies even under these sparse action and sparse reward constraints.

Our research advances agricultural sustainability by leveraging RL to address water scarcity while promoting environmental conservation. By introducing a scalable and adaptable framework for optimizing irrigation, our findings bridge the gap between theoretical

advancements and practical applications. These insights offer actionable strategies for policymakers and agricultural practitioners, particularly in water-scarce regions, enabling informed decision-making that aligns economic goals with environmental stewardship.

In the remainder of this paper, we review related work on RL in agriculture in Section 2. Section 3 details our methodological approach, including the integration of the Proximal Policy Optimization (PPO) algorithm with the AquaCrop-OSPy crop simulation model and the development of our innovative reward mechanism that balances water conservation with yield optimization. In Section 4, we present our experimental findings, demonstrating how the RL-based irrigation strategies outperform conventional methods in terms of water usage reduction and profitability enhancement. Finally, Section 5 summarizes our key insights and discusses potential avenues for future research to advance sustainable irrigation practices through reinforcement learning further.

2. Related Works

Reinforcement learning (RL) has emerged as a promising approach for addressing the complexities of sustainable crop production in agricultural management. Its ability to make sequential decisions and adapt to dynamic environments makes it particularly effective for optimizing agricultural practices influenced by unpredictable weather conditions and soil variability [15]. However, unlike RL applications in domains such as robotics, where actions are frequent and rewards are often immediate, agricultural scenarios inherently feature sparse actions and delayed rewards. Irrigation interventions may be required only intermittently, and their true impact on yield is not realized until the end of the growing season, posing unique challenges for both the agent's learning and policy evaluation.

2.1. RL in Crop Simulation Models

Recent advancements have explored the integration of RL with crop simulation models to enhance crop management strategies. For instance, Gautron et al. [19] presented Gym-DSSAT, an RL environment derived from the Decision Support System for Agrotechnology Transfer (DSSAT) [11], enabling the simulation of management tasks such as fertilization and irrigation. Initial results indicated that RL-based policies could surpass conventional expert-designed strategies, achieving higher yields with reduced resource input. Similarly, Turchetta et al. [16] developed CyclesGym, an RL environment based on the Cycles crop model [21], which optimizes long-term crop management tasks such as nitrogen fertilization and crop rotation. This approach balances yield optimization with environmental objectives, including mitigating nitrate leaching.

Kallenberg et al. [18] extended RL applications through CropGym, based on the PCSE crop model [30], focusing on optimizing nitrogen fertilization for winter wheat. The study demonstrated that RL could achieve near-optimal nitrogen application policies, effectively reducing environmental harm from nutrient runoff. Furthermore, Madondo et al. [20] combined RL with the Soil and Water Assessment Tool (SWAT) [31] in SWATGym, providing a platform to optimize irrigation and fertilization practices using real-time data. These studies underscore the potential of RL in agricultural management, though many focus on combined fertilization and irrigation strategies, complicating the isolation of irrigation-specific impacts.

2.2. RL Applications in Irrigation Management

Other studies have applied reinforcement learning (RL) directly to irrigation management using real-time data and machine learning models, often targeting specific crops. For example, Chen et al. [23] employed a deep Q-Network (DQN) algorithm to optimize irrigation schedules for rice fields, demonstrating significant water savings without nega-

tively affecting crop yield. Similarly, Alibabaei et al. [25] enhanced DQN models with Long Short-Term Memory (LSTM) networks to incorporate time-series data on soil conditions, leading to improvements in water use efficiency and productivity compared to fixed irrigation schedules. In the context of orchard crops, Ding and Du [24] developed DRLIC, a deep reinforcement learning (DRL)-based irrigation system for almond orchards. This system utilized real-time soil moisture data and incorporated safety mechanisms to prevent over- or under-irrigation, achieving reduced water usage while maintaining crop health.

Although these approaches demonstrate the potential of RL in optimizing water use for specific crops, their lack of integration with comprehensive crop simulation models limits their scalability and adaptability to diverse environmental conditions. By relying primarily on real-time data and crop-specific implementations, these methods may struggle to generalize across varied agricultural systems or account for long-term environmental factors.

2.3. AquaCrop-Gym

A significant step forward in RL for irrigation management was made by the authors of [26], who introduced AquaCrop-Gym. This framework integrates the AquaCrop-OSPy model [22,32] with OpenAI Gym to evaluate the effects of RL-driven irrigation strategies on crop yield. Their findings demonstrated that deep reinforcement learning (DRL) outperformed conventional heuristic methods under extreme conditions, such as zero rainfall or severe water restrictions. However, in scenarios with high rainfall variability, conventional soil moisture threshold (SMT) methods performed comparably or better.

AquaCrop-Gym highlights the potential of RL in optimizing irrigation strategies but also underscores critical limitations, including a reward system focused solely on profitability. These gaps provide an opportunity for further research to develop RL-driven frameworks that explicitly prioritize water efficiency while maintaining adaptability across diverse environmental contexts.

2.4. Research Gaps and Contributions

Despite significant progress in RL applications for crop and irrigation management, several critical research gaps remain. Many studies simultaneously address irrigation and fertilization, making it challenging to isolate the specific impacts of irrigation strategies on water efficiency. Existing RL reward mechanisms often emphasize either penalizing water use or maximizing yield independently, without integrating both objectives into a balanced framework. Moreover, while agriculture inherently involves sparse actions—interventions are not continuously needed—and rewards that only manifest at the end of the growing season, few studies have explicitly accounted for these intrinsic characteristics. This oversight may hinder the development of RL agents capable of learning effective long-term strategies.

This study addresses these gaps by focusing exclusively on irrigation optimization and introducing several key innovations. First, we propose a novel reward system that balances water conservation and yield maximization, incorporating step penalties for excessive irrigation and incentives for end-of-season yield. This design inherently accommodates the sparse action and delayed reward nature of agricultural RL tasks, encouraging agents to act only when necessary and strive for long-term performance. Second, we utilize historical weather data spanning 1982 to 2018 for Champion, Nebraska, as provided by the AquaCrop-OSPy model, ensuring authenticity and facilitating evaluation under realistic conditions. Third, AquaCrop-OSPy is integrated with Gymnasium, a stable and actively maintained successor to OpenAI Gym, improving compatibility with modern RL tools and ensuring long-term scalability. Finally, this study evaluates RL-driven strategies using

a comprehensive benchmarking framework, including SMT, rainfed, interval-based, net irrigation, and random strategies, providing a rigorous comparison to assess performance.

By addressing these gaps and tailoring RL approaches to the sparse, delayed nature of agricultural decision-making, this study advances RL applications in irrigation management. It offers actionable insights for real-world adoption and scalability, paving the way for more sustainable and efficient agricultural practices.

3. Materials and Methods

This study employs reinforcement learning (RL) to optimize irrigation scheduling for maize crops, aiming to enhance water efficiency without compromising yield. Unlike many RL applications, such as robotics, where actions and rewards are dense, agricultural irrigation management involves sparse actions (infrequent interventions) and sparse rewards (yield realized only at the end of the season). To effectively address these challenges, we leverage Proximal Policy Optimization (PPO) [29], a state-of-the-art RL algorithm known for its stability and efficiency in complex decision-making tasks.

Our approach compares PPO-based irrigation strategies to conventional methods that rely on soil moisture thresholds or fixed intervals, as well as two baseline scenarios: rainfed (no irrigation) and a random policy. To ensure a robust and fair comparison, we optimized both the PPO algorithm's hyperparameters and the soil moisture thresholds using the Optuna framework [33], which efficiently explores large hyperparameter spaces. This systematic tuning ensures that each method's performance is measured under near-optimal conditions, enhancing the validity and reliability of our results.

All experiments were conducted on the Aziz supercomputer at King Abdulaziz University, utilizing A100 GPUs with 32 CPU cores to ensure computational efficiency and scalability.

3.1. Reinforcement Learning Environment

Our work builds upon the `aquacrop-gym` environment introduced by Kelly et al. [26], which integrates the AquaCrop-OSPy model with OpenAI Gym. To improve compatibility and long-term maintainability, we transitioned to Gymnasium, the actively maintained successor to OpenAI Gym. Gymnasium offers better support for modern RL libraries, enhanced stability, and streamlined development, making it an ideal platform for implementing adaptive irrigation strategies.

In addition, we updated AquaCrop-OSPy from version 1.1.2 to version 3.0.9 to incorporate the latest features and improvements. Dependency management was handled using Poetry [34], simplifying the installation process and ensuring a reproducible environment. These modifications provide the RL agent with comprehensive access to crop and environmental parameters, enabling more informed and adaptive decisions.

We selected PPO from the Stable-Baselines3 library [35] due to its proven effectiveness in policy optimization tasks with discrete actions. PPO's ability to balance exploration and exploitation while maintaining training stability is critical for handling the sparse action and sparse reward nature of agricultural RL tasks, where daily irrigation decisions have long-term repercussions on yield and water efficiency.

To ensure seamless integration between decision-making, crop simulation, and reward evaluation, we structured the RL framework as an iterative process. At each timestep, the RL agent receives a 26-dimensional state vector comprising crop, soil, and weather parameters from AquaCrop-OSPy. Based on this information, the agent selects an irrigation action, which is then applied to the crop model. AquaCrop-OSPy updates the crop state by simulating water balance and biomass accumulation, returning the new state and a reward to the RL agent. The reward function dynamically balances water conservation and yield

maximization, allowing the agent to refine its policy through trial and error over multiple training episodes.

The integration of AquaCrop-OSPy with Gymnasium ensures that the RL agent learns adaptive irrigation strategies that are both data-driven and environmentally sustainable. The iterative interaction between observation, action, simulation, and reward assignment enables the agent to develop optimal irrigation schedules that minimize water usage while maximizing crop yield.

3.2. Simulation Setup and Data

We employed historical weather and soil data from AquaCrop-OSPy [32] for Champion, Nebraska, spanning 1982 to 2018. Each growing season starts on May 1, with the crop grown in Sandy Loam soil—known for its balanced water retention and drainage. The initial soil water content was set to Field Capacity to ensure optimal starting conditions. Weather data from 1982 to 2007 were used for training, and data from 2008 to 2018 for evaluation, providing a realistic temporal split and ensuring that the trained policies were tested on unseen conditions.

AquaCrop-OSPy simulates crop growth and water balance under varying irrigation regimes, providing a high-fidelity environment for RL training and evaluation. By combining authentic weather patterns, soil characteristics, and initial conditions, we ensured that the resulting irrigation policies are both scientifically sound and practically applicable.

3.3. State and Action Spaces

At each daily timestep, the RL agent receives a 26-dimensional observation vector encompassing crop parameters (e.g., age in days, canopy cover, biomass growth, soil water depletion, and total available water) and weather data (daily precipitation, minimum/maximum temperatures, and aggregated weather summaries over the preceding seven days). This rich observation space allows the agent to make context-aware decisions that consider both immediate soil moisture conditions and long-term yield implications.

The action space consists of two discrete options at each timestep: apply no irrigation (0 mm) or apply 25 mm of water. This selection is guided by agronomic best practices, ensuring practical applicability in real-world irrigation management. Fixed irrigation depths per event are widely used due to pump capacity, soil infiltration rates, and operational constraints. Studies by [36,37] support 25–35 mm irrigation depths as effective for maintaining soil moisture while ensuring high crop productivity. Steele et al. [36] found that applying 25 mm per irrigation event produced yields comparable to other strategies while optimizing water use efficiency. Similarly, Irmak et al. [37] confirmed that irrigation depths in this range sustain high water use efficiency without excessive percolation losses. Restricting to a binary choice reduces complexity and directs the agent's focus toward determining the optimal timing of interventions rather than the quantity of water to apply. Such simplicity is advantageous in real-world settings, where flexible but efficient strategies can easily integrate into existing irrigation systems.

Figure 1 illustrates the daily interaction loop between the PPO agent and the AquaCrop-OSPy simulation within the AquaCropGymnasium framework. At the start of each day, the agent receives an observation vector describing the crop state (e.g., canopy cover, biomass growth, and soil water depletion) and environmental factors (e.g., precipitation and temperature). Guided by this information, the agent selects an irrigation action (0 mm or 25 mm). AquaCrop-OSPy then simulates the crop and soil responses over the course of that day, returning updated conditions—such as soil moisture and water balance—to the agent at the next timestep. Through this iterative feedback process, the

agent refines its strategy, learning when and how much to irrigate to maximize yield while conserving water.
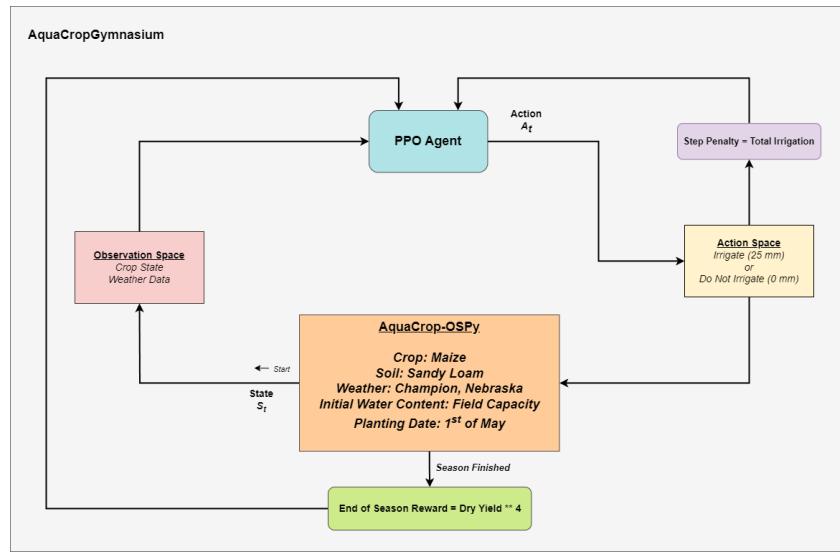

Figure 1. The interaction between the PPO agent and the AquaCrop-OSPy simulation within the AquaCropGymnasium framework. The notation ** represents exponentiation, where the end-of-season reward is calculated as the dry yield raised to the power of 4.

The reward mechanism combines step-based penalties for irrigation use with a final, yield-based reward at the end of the growing season. Each daily irrigation event incurs a penalty proportional to the amount of water applied, discouraging excessive water use. At the season's conclusion, the agent receives a terminal reward based on the final dry yield, thereby incentivizing it to maintain productivity. Unlike the purely profit-focused approach of Kelly et al. [26], our method explicitly integrates sustainability by balancing short-term resource use against long-term yield outcomes. Over successive growing seasons, the agent learns to adopt irrigation practices that enhance crop yields while minimizing water consumption.

3.4. Reward Mechanism

A well-designed reward function is crucial for guiding the RL agent toward strategies that balance water conservation and yield. In this study, the reward function consists of incremental penalties for irrigation events and a yield-based terminal reward, aligning with the sparse action and delayed reward characteristics inherent in agricultural tasks.

At each timestep t, the agent incurs a penalty based on the total cumulative irrigation applied so far:

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

where Penalty_t (mm) represents the penalty at timestep t. $\text{TotalIrrigation}_t = \sum_{k=1}^{t} I_k$ (mm) is the cumulative irrigation applied up to time t, reinforcing the impact of past watering events on current decisions. I_t (mm) is the irrigation depth applied at time t.

This cumulative penalty discourages excessive irrigation by dynamically increasing penalties for each watering event, reinforcing the importance of efficient scheduling. The penalty function is integrated into the AquaCrop-OSPy framework, ensuring penalties accumulate in real time as irrigation is applied.

At the end of the episode (the growing season), the agent receives a terminal reward based on the final crop yield output from the AquaCrop-OSPy model:

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

where DryYield (t/ha) is the final dry yield of the maize crop.

The final yield is obtained directly from the crop growth simulation, ensuring that reward feedback is grounded in realistic biophysical responses to irrigation. Raising the yield to the fourth power amplifies small yield differences, incentivizing precise irrigation strategies that optimize long-term productivity. Without this exponentiation, small differences in yield (e.g., 13.5 t/ha vs. 13.8 t/ha) might not significantly impact the agent's decisions, potentially leading to suboptimal irrigation policies.

The total episode reward is computed as:

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

where T represents the total number of timesteps in the growing season.

This reward formulation integrates both immediate and long-term incentives. The incremental penalty discourages the overuse of water, while the amplified terminal reward ensures that the agent remains focused on yield maximization rather than purely minimizing irrigation. Through iterative learning, the RL agent refines its irrigation policy to balance short-term resource conservation with long-term crop performance.

The reward function operates dynamically within the AquaCrop-OSPy framework. The penalty component is applied at each timestep, ensuring that excessive irrigation is continuously discouraged. The final reward, obtained from AquaCrop-OSPy's crop model, directly influences the agent's policy by reinforcing the long-term value of optimal irrigation scheduling. By combining these components, the RL agent develops sustainable irrigation strategies that enhance water use efficiency while maintaining high agricultural productivity.

3.5. Proximal Policy Optimization (PPO)

To develop an adaptive irrigation management system, we employ the Proximal Policy Optimization (PPO) algorithm [29], a state-of-the-art policy gradient method known for its robustness and sample efficiency in complex environments. PPO optimizes the policy by maximizing a clipped surrogate objective function:

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

where $r_t(\theta)$ is the probability ratio, $\hat{A}_t$ the advantage estimate, ϵ the clipping parameter, and θ the policy parameters. This objective function encourages beneficial policy updates while preventing drastic changes that could destabilize learning. Both the policy and value functions in PPO use neural networks with two hidden layers of 64 neurons each, employing ReLU activation functions. The policy network outputs a probability distribution over actions, while the value network estimates the expected return from a given state.

In this study, PPO was implemented using the Stable-Baselines3 library [35] and integrated into the AquaCrop-OSPy simulation through the Gymnasium framework. The RL agent interacts with the simulation environment, receiving daily observations of crop and environmental parameters (e.g., soil moisture and precipitation) and selecting irrigation actions (0 mm or 25 mm) at each timestep. The clipped surrogate objective in PPO ensures

stable learning in this sparse-reward setting by preventing overly aggressive policy updates, enabling the agent to refine its irrigation decisions over successive episodes.

This configuration ensures that PPO can handle the complexity and uncertainty of agricultural irrigation tasks, learning robust and scalable policies that promote water efficiency and high yields. By leveraging AquaCrop-OSPy's detailed crop–water interaction model, PPO dynamically optimizes irrigation timing, balancing water conservation and productivity under diverse environmental conditions.

3.6. PPO Hyperparameter Optimization

To ensure optimal performance, we used Optuna [33] to systematically tune the hyperparameters of the PPO agent. This framework efficiently explores large hyperparameter spaces, identifying configurations that strike a balance between exploration and exploitation.

We conducted 50 trials, each involving training the PPO agent for 100,000 timesteps with a unique hyperparameter set sampled from defined ranges (Table 1). In each trial, Optuna sampled a combination of hyperparameter values within the specified ranges, including learning rate, number of steps per update, batch size, number of epochs, discount factor, clip range, and entropy coefficient. These sampled configurations were then used to train the PPO agent on the reinforcement learning environment for 100,000 timesteps. After training, the agent's performance was evaluated by calculating the mean cumulative reward over the last ten episodes, which served as the objective function for the hyperparameter optimization process. Our objective was to maximize the mean cumulative reward over the last ten training episodes. By focusing on final performance, we ensured that the agent not only learned effectively but also generalized well by the end of training.

Final performance was assessed based on the mean cumulative reward achieved over the last ten training episodes in each trial. This metric was selected as it encapsulates the agent's ability to optimize irrigation schedules by balancing water conservation—through minimizing penalties for excessive irrigation—and yield maximization—through achieving high crop productivity. Evaluating performance during the final episodes allowed for a robust assessment of the agent's policy, ensuring that it could generalize effectively to diverse environmental conditions. This approach was particularly suited to addressing the sparse-action, sparse-reward structure inherent in agricultural irrigation tasks. This systematic methodology enabled the identification of hyperparameter configurations that effectively balanced learning stability, exploration, and exploitation.

Table 1. Hyperparameter search ranges for PPO optimization.

Hyperparameter	Search Range
Learning Rate	Log-uniform between 1×10^{-5} and 1×10^{-3}
Number of Steps per Update (`n_steps`)	{1024, 2048, 4096}
Batch Size	{128, 256, 512, 1024}
Number of Epochs	Integers from 10 to 50
Discount Factor (γ)	Uniform between 0.90 and 0.99
Clip Range (ϵ)	Uniform between 0.1 and 0.3
Entropy Coefficient	Log-uniform between 1×10^{-5} and 1×10^{-3}

After optimization, we identified the hyperparameters listed in Table 2 as yielding the best performance. These settings provided a stable training regime that effectively accounted for the sparse action and delayed reward structure of the agricultural domain.

Key hyperparameters include a relatively high discount factor ($\gamma = 0.98$) to emphasize long-term outcomes, and an entropy coefficient (4.50×10^{-4}) that encourages sufficient exploration. The selected clip range ($\epsilon = 0.22$) moderates policy updates, maintaining

training stability, while the chosen learning rate (6.34×10^{-4}) and batch size (512) offer a good balance between convergence speed and gradient estimate quality.

Table 2. Optimized PPO hyperparameters.

Hyperparameter	Value
Learning Rate (α)	6.34×10^{-4}
Number of Steps per Update (`n_steps`)	2048
Batch Size	512
Number of Epochs	23
Discount Factor (γ)	0.98
Clip Range (ϵ)	0.22
Entropy Coefficient	4.50×10^{-4}

With these optimal hyperparameters, we retrained the PPO agent for various durations (500,000, 1,000,000, 1,500,000, 2,000,000, and 2,500,000 timesteps) to assess how the agent's performance evolved over extended training periods. Evaluating policies saved at these checkpoints provided insights into the stability, consistency, and reliability of the learned irrigation strategies.

By integrating PPO with the optimized hyperparameters in our custom RL environment, we established a robust foundation for RL-based irrigation management. This setup enabled us to compare our RL-driven solutions against both optimized and conventional irrigation strategies, ultimately providing a comprehensive evaluation of the agent's real-world applicability.

3.7. Irrigation Strategies

To comprehensively evaluate the RL-based PPO agent, we compared its performance against both optimized and conventional irrigation methods.

3.7.1. Optimized Strategies

Alongside the PPO agent, we optimized the soil moisture threshold (SMT) irrigation strategy. SMT applies irrigation when soil moisture falls below predefined thresholds tailored to specific crop growth stages. Using Optuna, we conducted 50 trials to fine-tune thresholds for each growth stage. The resulting optimal moisture thresholds were 23.72%, 26.46%, 38.19%, and 50.11% of total available water (TAW) for emergence, canopy growth, maximum canopy, and canopy senescence, respectively.

To prevent overirrigation, we imposed a 300 mm seasonal irrigation cap. This ensures sufficient soil moisture for crop growth without excessive water use. By combining threshold optimization with an irrigation limit, we enhance water efficiency while maintaining productivity.

3.7.2. Conventional Strategies

In addition to the optimized SMT approach, we implemented several conventional irrigation strategies commonly used in agricultural practice:

- **Interval-Based Irrigation**: Irrigation is applied every seven days, regardless of soil conditions. This approach reflects traditional fixed-interval practices and provides a straightforward baseline for comparison.
- **Net Irrigation**: Daily additions of water maintain soil moisture above 70% TAW, ensuring consistently favorable conditions for crop growth.

- **Rainfed Irrigation**: Serving as a natural baseline, this strategy relies solely on precipitation, allowing us to measure the impact of supplemental irrigation on yield and water efficiency.
- **Random Agent**: The random agent selects actions (0 mm or 25 mm) uniformly at each timestep, ignoring environmental states. This nondeterministic baseline tests the PPO agent's resilience and effectiveness compared to random decision-making.

3.8. Evaluation Framework and Performance Metrics

We evaluated the PPO agent and the optimized SMT strategy within the AquaCrop-OSPy simulation environment, which accurately models crop growth and soil–water dynamics under varying conditions. To ensure robust and statistically meaningful results, each irrigation method was tested over 100 episodes, each representing a distinct growing season with unique weather and soil characteristics. Using 100 episodes enhances statistical significance and captures the variability encountered in real-world farming scenarios.

To address stochastic effects and initial condition variability, we employed three different random seeds for each experiment. Averaging results across these seeds further improves the reliability and robustness of our findings.

We assessed irrigation methods using key performance metrics:

- **Dry Yield (t/ha):** Final maize yield at the end of the season.
- **Total Irrigation (mm):** Total volume of irrigation water applied.
- **Water Efficiency (kg/ha/mm):** Yield produced per millimeter of irrigation water, indicating how effectively water is converted into biomass.
- **Profitability (USD):** Net economic gain, factoring in both crop yield and irrigation costs.

We report average values and standard deviations across the three random seeds and 100 episodes for each metric, ensuring a comprehensive and reliable performance assessment.

4. Results and Discussion

This section presents a comparative analysis of our RL-based irrigation strategy and conventional methods. We focus on how the reward mechanism guides policy development, ensuring that water conservation does not come at the expense of yield or profit.

4.1. PPO Training Progress and Performance at Different Timesteps

To understand how our agent's strategy evolves, we evaluated its performance at training milestones of 500,000, 1,000,000, 1,500,000, 2,000,000, and 2,500,000 timesteps. Figure 2 shows normalized rewards across these checkpoints, and Table 3 summarizes key metrics—mean yield, seasonal irrigation, profit, and water efficiency.

Initially, the agent prioritized yield, achieving 13.95–14.00 t/ha at 500,000–1,000,000 timesteps but consuming more water (227–241 mm). Although profitability and water efficiency improved slightly, water usage had not yet been optimized.

By 1,500,000 timesteps, the agent found a favorable balance: yield remained robust (13.80 t/ha), seasonal irrigation dropped to 179.83 mm, and water efficiency surged to 76.76 kg/ha/mm, resulting in the highest profit (USD 576.91). This milestone reflects the reward mechanism's impact, aligning economic returns with environmental sustainability.

Beyond 1,500,000 timesteps, further training improved water efficiency at the expense of yield and profit. At 2,000,000 timesteps, yields declined to 13.51 t/ha, profits fell to USD 538.27, and water efficiency reached 81.33 kg/ha/mm. By 2,500,000 timesteps, yield (13.64 t/ha) and profit (USD 560.41) partially recovered while efficiency reached

81.72 kg/ha/mm. These fluctuations indicate that excessive emphasis on water savings can diminish overall economic viability.

In essence, 1,500,000 timesteps emerged as the "sweet spot" where yield, profitability, and water efficiency were optimally balanced. This result underlines the importance of careful training duration selection, ensuring that the RL approach leverages the reward mechanism effectively to achieve sustainable and economically viable irrigation strategies.

In Section 4.2, we compare the best-performing RL strategy with conventional soil moisture threshold and interval-based methods. This analysis further illustrates how adapting irrigation decisions in response to real-time feedback—rather than relying on static thresholds—can promote more efficient resource use, stable yields, and stronger economic returns.

(**a**) 500,000 Timesteps

(**b**) 1,000,000 Timesteps

(**c**) 1,500,000 Timesteps

(**d**) 2,000,000 Timesteps

(**e**) 2,500,000 Timesteps
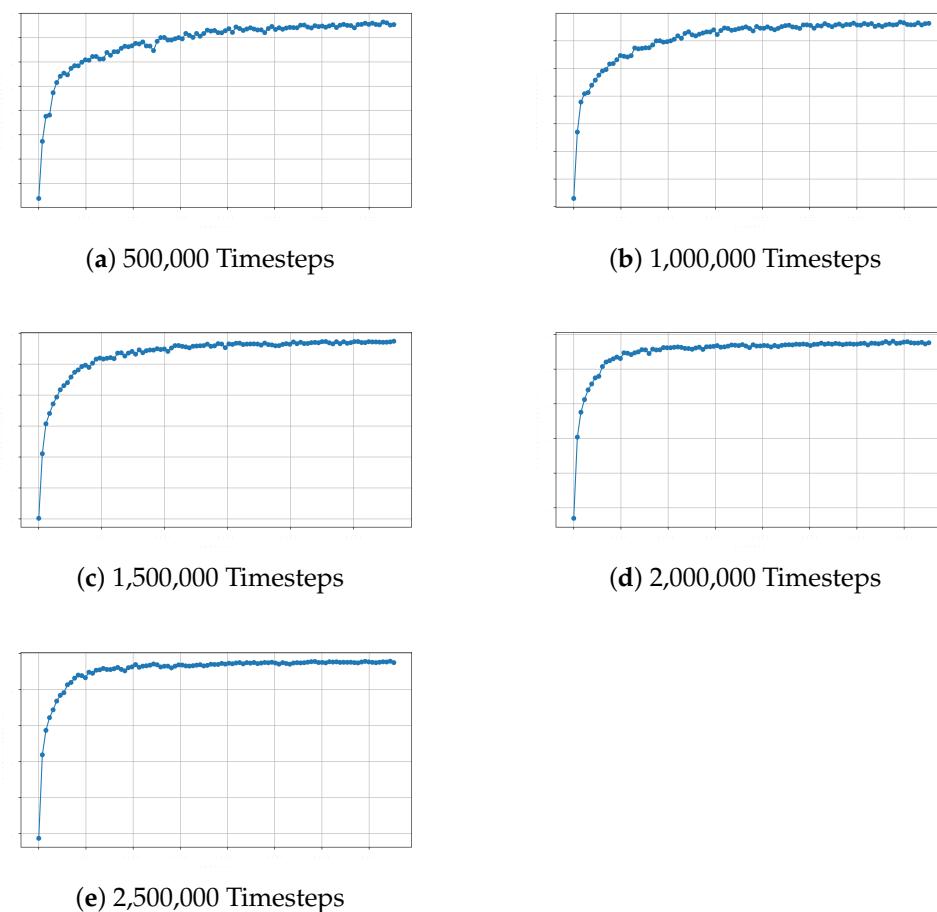
Figure 2. Normalized rewards per episode for the PPO agent at different training milestones (500 K to 2.5 M timesteps). The overall upward trend indicates effective policy learning and performance optimization over time.

Table 3. PPO agent performance metrics at various training stages.

Training Steps	Mean Yield (tonne/ha)	Seasonal Irrigation (mm)	Profit (USD)	Water Efficiency (kg/ha/mm)
500,000	14.00 ± 0.22	240.75 ± 86.59	551.10	58.15
1,000,000	13.95 ± 0.21	226.92 ± 88.15	556.81	61.49
1,500,000	13.80 ± 0.17	179.83 ± 89.44	576.91	76.76
2,000,000	13.51 ± 0.16	166.17 ± 85.95	538.27	81.33
2,500,000	13.64 ± 0.28	166.92 ± 87.26	560.41	81.72

4.2. Comparison of Irrigation Strategies

Table 4 compares the key performance metrics (yield, irrigation volume, water use efficiency, and profitability) for the PPO strategy and several conventional irrigation methods. The PPO approach stands out by achieving significantly better water conservation, higher water efficiency, and improved profitability, while maintaining yield levels comparable to other top-performing strategies. The optimized soil moisture threshold (SMT) method ranks second, demonstrating strong overall performance though slightly less efficient in water use than PPO. In contrast, while the random irrigation approach attains the highest absolute yield, it does so at the cost of excessive water usage and substantial financial losses.

Table 4. Comparative performance of PPO and conventional irrigation strategies.

Strategy	Yield (t/ha)	Irrigation (mm)	Water Efficiency (kg/ha/mm)	Profit (USD/ha)
PPO	13.80 ± 0.17	179.83 ± 89.44	76.76	576.91
Thresholds (Optimized)	13.95 ± 0.20	255.00 ± 112.05	54.72	528.83
Net Irrigation	13.98 ± 0.23	305.61 ± 119.17	45.73	482.21
Interval-Based	13.81 ± 0.70	380.17 ± 33.69	36.32	377.42
Random	14.02 ± 0.24	1640.83 ± 141.89	8.54	−845.60
Rainfed	8.88 ± 3.75	0.0	N/A	−130.19

4.2.1. Crop Yield (t/ha)

Crop yield indicates the productivity of each irrigation strategy. As shown in Figure 3, the random strategy achieves the highest yield (14.02 t/ha), followed closely by net irrigation (13.98 t/ha) and the optimized SMT approach (13.95 t/ha). The PPO agent, at 13.80 t/ha, delivers a slightly lower yield than these methods, but the differences are marginal. All four approaches—random, net, SMT, and PPO—demonstrate high overall productivity.
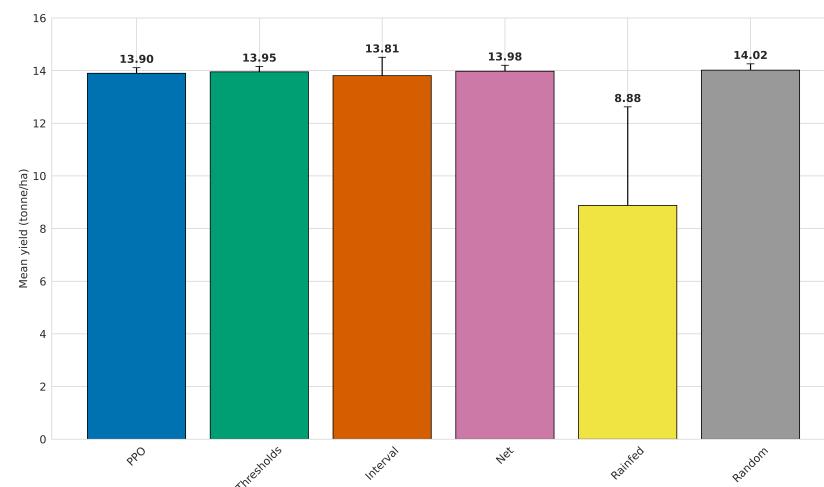

Figure 3. Maize yields (t/ha) under different irrigation strategies. Although the random strategy achieves the highest yield, its excessive water use and resulting financial losses limit its practicality. The PPO, SMT, and net irrigation methods provide strong yields with more sustainable water use. Rainfed conditions yield significantly less production due to reliance on natural rainfall alone.

While PPO does not achieve the absolute highest yield, it maintains a strong balance between yield and resource efficiency, outperforming other strategies in terms of water use efficiency and profitability. Although SMT uses slightly more water than PPO, it still offers a substantial improvement in efficiency and productivity over more traditional methods, making it a viable option for farmers who prefer simpler threshold-based approaches.

The rainfed strategy, relying solely on natural precipitation, predictably results in the lowest yield (8.88 t/ha). Despite its minimal water input, its inconsistent and insufficient moisture supply hinders productivity compared to strategies that supplement rainfall with irrigation.

It is worth noting that the AquaCrop-OSPy simulation may not fully capture the negative impacts of severe overirrigation, as evidenced by the random agent's strong yields despite excessive water use. Real-world scenarios would likely see diminished returns from overirrigation due to factors like waterlogging and nutrient leaching [38]. This suggests that models and reward structures should incorporate more realistic penalties for overirrigation to better align simulated outcomes with actual on-farm conditions.

Overall, the PPO strategy stands out as a sustainable and economically beneficial approach, delivering strong yields without the excessive water usage seen in the random strategy or the less adaptive thresholds of SMT. This balance makes PPO particularly valuable in contexts where water resources are limited and both productivity and sustainability are top priorities.

4.2.2. Irrigation (mm)

Irrigation volume, measured in millimeters (mm), reflects the total amount of water applied during the growing season. This metric indicates how efficiently each strategy meets the crop's water demands without wasteful over-application.

As shown in Figure 4, the PPO strategy applies 179.83 mm of water, achieving the highest water-use efficiency while sustaining robust yields. This amount is approximately 30% less than the optimized soil moisture threshold (SMT) strategy's usage (255.00 mm), underscoring PPO's precision in the timing and quantity of irrigation. By providing water only when needed, PPO prevents both over- and under-irrigation, thus maximizing benefits in yield and resource conservation.
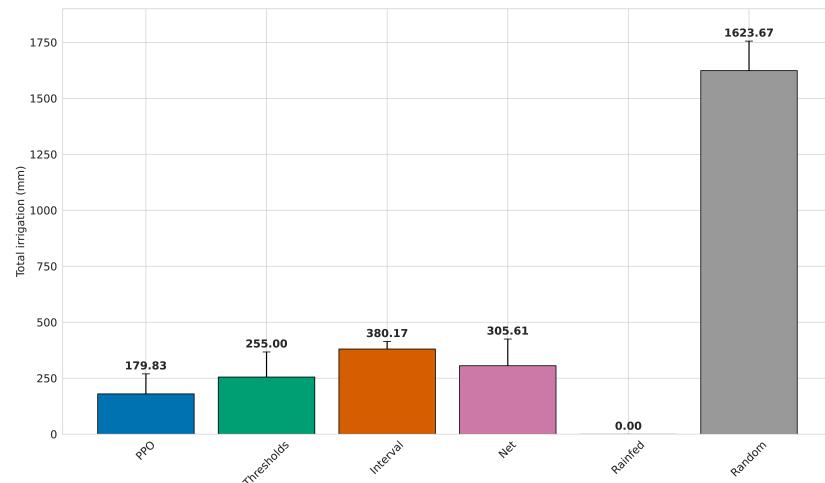

Figure 4. Total irrigation applied (mm) under different strategies. PPO and SMT demonstrate efficient water use, while random irrigation applies excessive amounts. Rainfed conditions rely solely on natural precipitation.

The SMT method, although using more water than PPO, still outperforms conventional practices. It conserves 33% more water than interval-based irrigation and 17% more than net irrigation by responding only when soil moisture levels drop below set thresholds. This adaptive approach ensures sufficient moisture for optimal growth without unnecessary application.

In contrast, the random strategy applies a staggering 1640 mm of water—far exceeding the crop's needs. While this overabundance leads to the highest yield, it also results in enormous resource waste and significant operational costs, illustrating the drawbacks of indiscriminate irrigation.

Rainfed conditions, relying solely on natural rainfall, require no supplemental irrigation. Although this approach conserves the most water, its unpredictable and often inadequate moisture supply leads to substantially lower yields.

4.2.3. Water Efficiency (kg/ha/mm)

Water efficiency, measured in kilograms of yield per hectare per millimeter of irrigation water, indicates how effectively each strategy converts water into crop production.

Figure 5 reveals that PPO achieves the highest water efficiency at 76.76 kg/ha/mm, outperforming all other methods. This exceptional efficiency—40% greater than that of the optimized SMT strategy (54.72 kg/ha/mm)—demonstrates PPO's capacity to maintain strong yields while minimizing water inputs. PPO's precision in irrigation timing and volume ensures that each unit of water contributes maximally to crop growth, making it the most sustainable choice, especially in water-scarce environments.
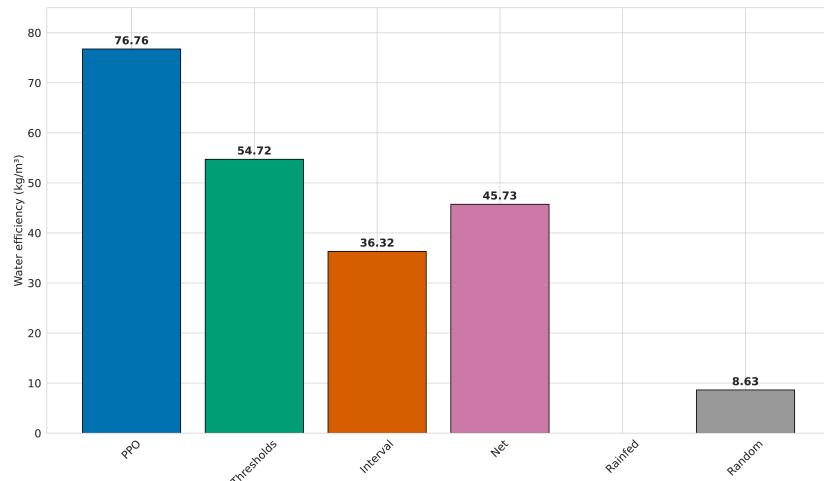

Figure 5. Water efficiency (kg/ha/mm) achieved under various irrigation strategies. PPO demonstrates the highest efficiency, followed by SMT. Rainfed conditions do not apply since no additional water is used.

While SMT is less efficient than PPO, it remains the top-performing conventional strategy, offering significantly better efficiency than interval-based or net irrigation methods. By applying water only when soil moisture levels fall below defined thresholds, SMT reduces wastage and enhances productivity.

PPO's superior water efficiency and consistently high yields make it an ideal strategy for regions facing limited water availability. Its ability to deliver strong economic and agronomic performance, while minimizing environmental impact, underscores its value as a key tool for sustainable and productive agricultural systems.

4.2.4. Profit (USD/ha)

Profitability measures the net economic return per hectare after deducting water costs and other operational expenses. Consistent with Kelly et al. [26], we assume a crop price of USD 180/tonne, an irrigation cost of USD 1/ha-mm, and fixed non-irrigation production costs of USD 1728/ha. Under these conditions, our PPO strategy emerges as the most profitable, reflecting its superior balance between yield and water use efficiency.

In comparing our results to prior work, we focus specifically on studies using SMT and PPO. Kelly et al. [26] reported profits of USD 521.00/ha for SMT and USD 520.10/ha for PPO under similar cost assumptions. Our SMT approach achieves USD 528.83/ha, slightly exceeding their SMT profit, indicating incremental gains through refined threshold management. More notably, our PPO strategy attains USD 576.91/ha, surpassing their PPO profit by approximately 11%, demonstrating that a reward mechanism emphasizing both yield and water conservation, paired with a daily decision interval, can significantly enhance profitability (Figure 6).
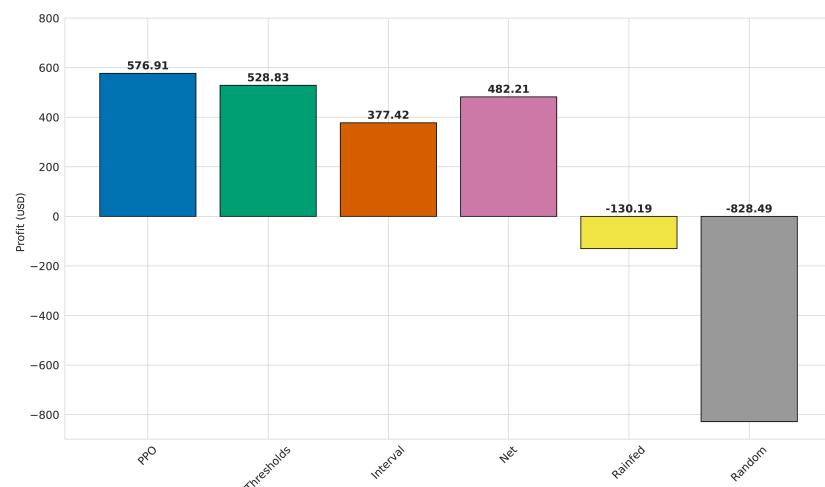
Figure 6. Profit per hectare (USD/ha) under different irrigation strategies. PPO achieves the highest profitability, surpassing the optimized SMT strategy by approximately 9%. Random and rainfed approaches yield negative profits (−) due to inefficient water use or reduced yields.

Table 5 highlights these differences. While both SMT and PPO strategies reported here outperform those from Kelly et al. [26], the PPO approach shows the most substantial improvement. The key differentiators are the integration of water efficiency into the reward function and the daily decision-making interval, which together foster a more adaptive and resource-conscious irrigation policy. This synergy of factors ultimately yields a more sustainable and profitable irrigation management solution for regions challenged by limited water availability.

Table 5. Comparison of profit (USD/ha) for SMT and PPO strategies between this study and Kelly et al. [26].

Study	SMT Profit (USD/ha)	PPO Profit (USD/ha)
Kelly et al. [26]	521.00	520.10
This Study	528.83	576.91

4.3. Long-Term Environmental Benefits

The implementation of our PPO-based RL irrigation strategy shows strong potential for delivering meaningful long-term environmental benefits. By reducing irrigation water usage by approximately 29% compared to the optimized soil moisture threshold (SMT) method (see Table 4), our approach can help alleviate pressures on groundwater reserves, addressing a critical concern where agriculture heavily contributes to groundwater depletion [3].

Additionally, the PPO strategy achieves the highest water efficiency (76.76 kg/ha/mm), surpassing conventional irrigation techniques (see Table 4 and Figure 5). This improved efficiency could, over time, lead to more sustainable water usage, helping reduce nutrient

runoff and soil erosion. By ensuring that irrigation occurs only when necessary, the PPO model minimizes ecological disturbances, supports healthier ecosystems, and preserves natural habitats.

Moreover, our binary action framework—choosing simply whether or not to irrigate—helps prevent over-irrigation. This restraint not only cuts down on water waste but also lowers energy consumption associated with pumping and distribution. Such energy savings can decrease greenhouse gas emissions, further reducing the agricultural sector's carbon footprint and contributing to climate change mitigation efforts.

Taken together, these environmental advantages suggest that PPO-based RL irrigation strategies can steer agricultural practices toward greater sustainability. By optimizing water use, preserving vital resources, and limiting environmental harm, this approach offers a promising pathway for future agricultural resilience and ecological stewardship.

4.4. Bridging Simulation to Practical Implementation

While our results showcase the effectiveness of RL-based irrigation strategies in a simulated setting, implementing these solutions in real-world farming contexts requires careful consideration of existing infrastructure and technology.

Our PPO-based strategies, developed using the `aquacropgymnasium` environment, can readily integrate with modern irrigation systems—such as drip or sprinkler setups—through Internet of Things (IoT) platforms. These platforms enable real-time data collection from soil moisture sensors, weather stations, and other agricultural sensors, providing continuous feedback to the RL agent. By leveraging this data, the agent can make informed and timely irrigation decisions that align closely with environmental conditions and the underlying simulation assumptions. However, practical deployment must address critical factors such as sensor calibration, data transmission reliability, and network connectivity to ensure robust and effective operation under diverse and unpredictable field conditions.

The binary nature of our action space—irrigate or not—simplifies the process of automating irrigation practices. Unlike approaches that require precise control of irrigation depth, this straightforward decision-making structure aligns well with existing irrigation infrastructure, reducing the complexity of implementation. This simplicity also minimizes potential errors, making RL-driven systems more accessible to farmers who may lack technical expertise. Additionally, automation through IoT-based systems ensures real-time responsiveness, even in environments with limited resources, enhancing both water and energy efficiency.

To ensure successful adoption, the integration of RL-based strategies into existing farm management software is essential. User-friendly dashboards can present actionable insights such as historical and predicted irrigation schedules, water efficiency metrics, and crop performance trends. These interfaces, mirroring the data handling and decision-making processes of the simulation environment, provide farmers with intuitive tools to monitor and control irrigation practices. Educational and training programs tailored for farmers are also critical to demystifying the use of RL-driven systems and fostering confidence in their deployment.

Scalability remains a key consideration for real-world adoption. Practical challenges such as the cost of deploying IoT sensors, compatibility with heterogeneous irrigation systems, and ensuring data privacy must be addressed. Collaborative efforts between researchers, technology providers, and policymakers can help create affordable and adaptable solutions for diverse farming contexts. Moreover, pilot programs that test RL-based systems in small-scale farms can serve as proof-of-concept studies, building trust and encouraging broader adoption within the agricultural community.

By addressing these considerations, RL-based irrigation strategies can effectively transition from simulation to practice, offering a pathway toward more efficient, sustainable, and resilient agricultural water management.

5. Conclusions

This study demonstrates the effectiveness of reinforcement learning (RL), specifically the Proximal Policy Optimization (PPO) algorithm, in optimizing irrigation strategies for maize crops under conditions characterized by sparse actions and delayed (end-of-season) rewards. By integrating RL with AquaCrop-OSPy simulations within the Gymnasium framework, we developed an irrigation policy that successfully balances water use efficiency, crop yield, and profitability. Compared to an optimized soil moisture threshold (SMT) approach, PPO reduced seasonal water usage by approximately 29% without compromising yields and increased profitability by about 9%. These results establish PPO as both financially and environmentally advantageous.

A key innovation in this study lies in the reward mechanism, which penalizes excessive irrigation while rewarding end-of-season yields. Unlike prior research that often prioritizes either yield or profit, our approach explicitly incorporates water conservation as a fundamental objective. Consequently, PPO achieved a water efficiency of 76.76 kg/ha/mm, a 40% improvement over the SMT strategy's 54.72 kg/ha/mm. This substantial gain reflects the value of integrating environmental sustainability into the RL optimization process, making the approach particularly suitable for regions facing water scarcity.

The findings confirm that RL-driven irrigation management can address the intrinsic challenges posed by sparse actions and delayed rewards in agriculture. By dynamically adjusting irrigation schedules in response to environmental conditions and by embedding water efficiency goals into the learning process, the PPO strategy offers a scalable, adaptable, and resource-conscious solution. In turn, this approach supports more resilient agricultural practices that protect vital water resources and reduce environmental impact, all while sustaining strong economic returns.

Looking ahead, future research should explore the applicability of this RL framework to a variety of crops and agricultural contexts. Incorporating real-time environmental data and weather forecasts into the simulation process may further enhance adaptability and responsiveness. Additionally, conducting field trials would help validate the simulated outcomes, bridging the gap between theoretical models and on-farm implementations. Ultimately, the continued development and refinement of RL-based irrigation strategies have the potential to transform water management in agriculture, contributing to global food security and sustainability.

References

1. Food and Agriculture Organization of the United Nations. *The Future of Food and Agriculture: Trends and Challenges*; Technical report; Food and Agriculture Organization of the United Nations: Rome, Italy, 2017.

2. Kılıç, O.; Boz, I.; Eryılmaz, G.A. Comparison of conventional and good agricultural practices farms: A socio-economic and technical perspective. *J. Clean. Prod.* **2020**, *258*, 120666. [CrossRef]

3. Siebert, S.; Burke, J.; Faures, J.M.; Frenken, K.; Hoogeveen, J.; Döll, P.; Portmann, F.T. Groundwater use for irrigation—A global inventory. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 1863–1880. [CrossRef]

4. Davis, K.F.; Rulli, M.C.; Seveso, A.; D'Odorico, P. Increased food production and reduced water use through optimized crop distribution. *Nat. Geosci.* **2017**, *10*, 919–924. [CrossRef]

5. Biswas, A.K.; Tortajada, C. Groundwater: An unseen, overused and unappreciated resource. *Int. J. Water Resour. Dev.* **2024**, *40*, 1–6. [CrossRef]

6. Zhao, H.; Di, L.; Guo, L.; Zhang, C.; Lin, L. An Automated Data-Driven Irrigation Scheduling Approach Using Model Simulated Soil Moisture and Evapotranspiration. *Sustainability* **2023**, *15*, 12908. [CrossRef]

7. Kelly, T.; Foster, T.; Schultz, D.M. Assessing the value of adapting irrigation strategies within the season. *Agric. Water Manag.* **2023**, *275*, 107986. [CrossRef]

8. Pretty, J.; Benton, T.G.; Bharucha, Z.P.; Dicks, L.V.; Flora, C.B.; Godfray, H.C.J.; Goulson, D.; Hartley, S.; Lampkin, N.; Morris, C.; et al. Global assessment of agricultural system redesign for sustainable intensification. *Nat. Sustain.* **2018**, *1*, 441–446. [CrossRef]

9. Li, M.; Xu, Y.; Fu, Q.; Singh, V.P.; Liu, D.; Li, T. Efficient irrigation water allocation and its impact on agricultural sustainability and water scarcity under uncertainty. *J. Hydrol.* **2020**, *586*, 124888. [CrossRef]

10. Holzworth, D.; Huth, N.I.; Fainges, J.; Brown, H.; Zurcher, E.; Cichota, R.; Verrall, S.; Herrmann, N.I.; Zheng, B.; Snow, V. APSIM Next Generation: Overcoming challenges in modernising a farming systems model. *Environ. Model. Softw.* **2018**, *103*, 43–51. [CrossRef]

11. Hoogenboom, G.; Porter, C.H.; Boote, K.J.; Shelia, V.; Wilkens, P.W.; Singh, U.; White, J.W.; Asseng, S.; Lizaso, J.I.; Moreno, L.P.; et al. The DSSAT crop modeling ecosystem. In *Advances in Crop Modelling for a Sustainable Agriculture*; Burleigh Dodds Science Publishing: Sawston, UK , 2019; pp. 173–216.

12. Steduto, P.; Hsiao, T.C.; Raes, D.; Fereres, E. AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles. *Agron. J.* **2009**, *101*, 426–437. [CrossRef]

13. Bhatia, A.K. Crop Growth Simulation Modeling. In *Modelling and Simulation of Diffusive Processes: Methods and Applications*; Basu, S., Kumar, N., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 315–332. [CrossRef]

14. Shaikh, T.A.; Rasool, T.; Lone, F.R. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Comput. Electron. Agric.* **2022**, *198*, 107119. [CrossRef]

15. Gautron, R.; Maillard, O.A.; Preux, P.; Corbeels, M.; Sabbadin, R. Reinforcement learning for crop management support: Review, prospects and challenges. *Comput. Electron. Agric.* **2022**, *200*, 107182. [CrossRef]

16. Turchetta, M.; Corinzia, L.; Sussex, S.; Burton, A.; Herrera, J.; Athanasiadis, I.; Buhmann, J.M.; Krause, A. Learning Long-Term Crop Management Strategies with CyclesGym. In Proceedings of the Advances in Neural Information Processing Systems, New Orleans, LA, USA, 28 November–9 December 2022; Volume 35, pp. 11396–11409.

17. Overweg, H.; Berghuijs, H.N.; Athanasiadis, I.N. CropGym: A reinforcement learning environment for crop management. *arXiv* **2021**, arXiv:2104.04326. [CrossRef]

18. Kallenberg, M.G.; Overweg, H.; van Bree, R.; Athanasiadis, I.N. Nitrogen management with reinforcement learning and crop growth models. *Environ. Data Sci.* **2023**, *2*, e34. [CrossRef]

19. Gautron, R.; Padrón, E.J.; Preux, P.; Bigot, J.; Maillard, O.A.; Emukpere, D. gym-DSSAT: A crop model turned into a Reinforcement Learning environment. *arXiv* **2022**, arXiv:2207.03270. [CrossRef]

20. Madondo, M.; Azmat, M.; Dipietro, K.; Horesh, R.; Jacobs, M.; Bawa, A.; Srinivasan, R.; O'Donncha, F. A swat-based reinforcement learning framework for crop management. *arXiv* **2023**, arXiv:2302.04988.

21. Kemanian, A.R.; Shi, Y.; White, C.M.; Montes, F.; Stöckle, C.O.; Huggins, D.R.; Cangiano, M.L.; Stefani-Faé, G.; Nydegger Rozum, R.K. The Cycles Agroecosystem Model: Fundamentals, Testing, and Applications. 2023. Available online: https://ssrn.com/abstract=4188402 (accessed on 7 February 2025).

22. Kelly, T.; Foster, T. AquaCrop-OSPy: Bridging the gap between research and practice in crop-water modeling. *Agric. Water Manag.* **2021**, *254*, 106976. [CrossRef]

23. Chen, M.; Cui, Y.; Wang, X.; Xie, H.; Liu, F.; Luo, T.; Zheng, S.; Luo, Y. A reinforcement learning approach to irrigation decision-making for rice using weather forecasts. *Agric. Water Manag.* **2021**, *250*, 106838. [CrossRef]

24. Ding, X.; Du, W. Drlic: Deep reinforcement learning for irrigation control. In Proceedings of the 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Milano, Italy, 4–6 May 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 41–53. [CrossRef]

25. Alibabaei, K.; Gaspar, P.D.; Assunção, E.; Alirezazadeh, S.; Lima, T.M. Irrigation optimization with a deep reinforcement learning model: Case study on a site in Portugal. *Agric. Water Manag.* **2022**, *263*, 107480. [CrossRef]

26. Kelly, T.; Foster, T.; Schultz, D.M. Assessing the value of deep reinforcement learning for irrigation scheduling. *Smart Agric. Technol.* **2024**, *7*, 100403. [CrossRef]

27. Towers, M.; Terry, J.K.; Kwiatkowski, A.; Balis, J.U.; Cola, G.d.; Deleu, T.; Goulão, M.; Kallinteris, A.; KG, A.; Krimmel, M.; et al. Gymnasium. *Zenodo* **2023**. [CrossRef]

28. Brockman, G. OpenAI Gym. *arXiv* **2016**, arXiv:1606.01540.

29. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347. [CrossRef]

30. De Wit, A.; Boogaard, H.; Fumagalli, D.; Janssen, S.; Knapen, R.; van Kraalingen, D.; Supit, I.; van der Wijngaart, R.; van Diepen, K. 25 years of the WOFOST cropping systems model. *Agric. Syst.* **2019**, *168*, 154–167. [CrossRef]

31. Bieger, K.; Arnold, J.G.; Rathjens, H.; White, M.J.; Bosch, D.D.; Allen, P.M.; Volk, M.; Srinivasan, R. Introduction to SWAT+, a completely restructured version of the soil and water assessment tool. *JAWRA J. Am. Water Resour. Assoc.* **2017**, *53*, 115–130. [CrossRef]

32. Foster, T.; Brozović, N.; Butler, A.; Neale, C.; Raes, D.; Steduto, P.; Fereres, E.; Hsiao, T.C. AquaCrop-OS: An open source version of FAO's crop water productivity model. *Agric. Water Manag.* **2017**, *181*, 18–22. [CrossRef]

33. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Anchorage, AK, USA, 4–8 August 2019; pp. 2623–2631. [CrossRef]

34. Eustace, S.; Contributors, T.P. Poetry: Python Packaging and Dependency Management Made Easy. Software; Version 3.09. 2022. Available online: https://python-poetry.org (accessed on 7 February 2025).

35. Raffin, A.; Hill, A.; Gleave, A.; Kanervisto, A.; Ernestus, M.; Dormann, N. Stable-baselines3: Reliable reinforcement learning implementations. *J. Mach. Learn. Res.* **2021**, *22*, 1–8.

36. Steele, D.D.; Stegman, E.C.; Knighton, R.E. Irrigation management for corn in the northern Great Plains, USA. *Irrig. Sci.* **2000**, *19*, 107–114. [CrossRef]

37. Irmak, S.; Burgert, M.J.; Yang, H.; Cassman, K.G.; Walters, D.T.; Rathje, W.R.; Payero, J.O.; Grassini, P.; Kuzila, M.S.; Brunkhorst, K.J.; et al. Large-scale on-farm implementation of soil moisture-based irrigation management strategies for increasing maize water productivity. *Trans. ASABE* **2012**, *55*, 881–894. [CrossRef]

38. Song, L.; Jin, J.; He, J. Effects of severe water stress on maize growth processes in the field. *Sustainability* **2019**, *11*, 5086. [CrossRef]