



OPEN ACCESS

EDITED BY

Bijayalaxmi Mohanty,
National University of Singapore, Singapore

REVIEWED BY

Sohail Abbas,
Henan University, China
Nafees Akhter Farooqui,
Integral University, India
Imran Ali Lakhiar,
Jiangsu University, China

*CORRESPONDENCE

Ruipeng Tang
✉ 2057874@siswa.um.edu.my

RECEIVED 21 May 2025

ACCEPTED 13 October 2025

PUBLISHED 06 November 2025

CITATION

Tang R, Tang J, Abu Talip MS, Aridas NK and Guan B (2025) Reinforcement learning control method for greenhouse vegetable irrigation driven by dynamic clipping and negative incentive mechanism. *Front. Plant Sci.* 16:1632431. doi: 10.3389/fpls.2025.1632431

COPYRIGHT

© 2025 Tang, Tang, Abu Talip, Aridas and Guan. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Reinforcement learning control method for greenhouse vegetable irrigation driven by dynamic clipping and negative incentive mechanism

Ruipeng Tang^{1*}, Jianxun Tang², Mohamad Sofian Abu Talip¹, Narendra Kumar Aridas¹ and Binghong Guan³

¹Department of Electrical Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur, Malaysia, ²Faculty of Electronics and Electrical Engineering, Zhaoqing University, Zhaoqing, Guangdong, China, ³Faculty of Business and Economics, University of Malaya, Kuala Lumpur, Malaysia

Greenhouse vegetable production was a complex agricultural system influenced by multiple interrelated environmental and management factors. Its irrigation control was a critical but not singularly decisive component. Traditional irrigation methods often caused the water wastage, uneven resource utilization and limited adaptability to dynamic environmental conditions, thereby hindering the sustainable production efficiency. To address these challenges comprehensively, this study proposed an advanced irrigation control method by utilizing the enhanced reinforcement learning approach. The Enhanced Negative-incentive Proximal Policy Optimization (ENPPO) algorithm is introduced, which integrates the dynamic clipping functions and negative incentives to manage the intricacies of continuous action spaces and high-dimensional environmental states. By incorporating real-time sensor data and historical irrigation records, the ENPPO algorithm accurately predicts the optimal irrigation volumes aligned with various vegetable growth stages. Experimental results showed that ENPPO algorithm outperforms conventional methods such as PPO and TRPO in prediction accuracy, convergence efficiency and water resource utilization. It minimized both excessive and insufficient irrigation scenarios, thus promoting enhanced vegetable yield and quality while simultaneously reducing agricultural production costs. Overall, this study presented the versatile technical solution for intelligent irrigation management within greenhouse systems, highlighting its substantial potential to advance sustainable agricultural practices.

KEYWORDS

irrigation prediction method, greenhouse vegetable irrigation, reinforcement learning algorithm, sustainable agricultural development, greenhouse vegetable production

1 Introduction

Nowadays, most greenhouse vegetables in the northern regions of Vietnam and Thailand primarily rely on irrigation methods such as drip irrigation, sprinkler systems and soil moisture-based quantitative irrigation. Although these techniques are widely recognized as modern methods, suboptimal management and inadequate implementation often result in the water wastage and uneven resource utilization. The significant disparities in technological advancement and operational management between farms exacerbate these issues. The excessive irrigation was combined with the poor water quality characterized by high concentrations of sodium, chloride, bicarbonates and inadequate drainage, which contributes to the soil salinization and waterlogging, so the conditions negatively affect the vegetable growth and significantly reduce the overall crop quality (Chi et al., 2021; Wang et al., 2022). Additionally, the insufficient scientific knowledge among farmers regarding the optimal irrigation practices frequently leads to inappropriate water applications, which disregards the essential factors such as the specific soil characteristics, crop water requirements and local environmental conditions.

Moreover, the excessive or poorly managed irrigation, especially in areas with slopes and vulnerable soil types, can lead to severe soil erosion and water pollution. The runoff from such irrigated lands, often contaminated with residual pesticides and chemical fertilizers due to the excessive application practices, poses substantial risks to local water bodies and ecosystems (Farooq et al., 2021; Parkash et al., 2021). Considering these challenges, the current irrigation practices often fall short of effectively matching precise irrigation volumes with dynamic crop growth requirements and rapidly changing environmental conditions. To address these gaps, this study aims to develop and rigorously validate a reinforcement-learning-based irrigation control framework that (i) integrates real-time environmental sensing with crop growth state, (ii) adaptively optimizes irrigation volumes under non-stationary conditions, and (iii) quantifies gains in water-use efficiency, yield/quality, and operating cost versus conventional rule- or threshold-based methods. By leveraging online learning and adaptive decision policies, the proposed approach targets precise, context-aware irrigation that better matches dynamic crop requirements and rapidly changing environments while reducing waste and environmental externalities.

2 Literature review

In order to solve the above problems, some scholars had combined the intelligent irrigation technology and achieved some achievements. The simulation-based approaches have been employed to model agroecosystem interactions and optimize irrigation schedules under variable climate conditions (Tolomio and Casa, 2020). Machine learning techniques, particularly deep learning models such as long short-term memory (LSTM) networks, have been used to predict soil moisture content and irrigation needs

across different soil types and crop categories (Kashyap et al., 2021). Remote sensing methods utilizing OPTRAM and satellite data have allowed for large-scale monitoring of irrigated areas (Yao et al., 2022), while sensor-driven fuzzy logic systems integrated with Arduino platforms have demonstrated effective irrigation control for crops like tomatoes and chili (Singh et al., 2022). Additional developments include edge computing-based fertigation systems (Tran et al., 2023), rainwater pipe irrigation techniques (Marimuthu et al., 2024), fuzzy irrigation models for subtropical orchards (Xie et al., 2022), and decision support systems such as IrrigaSys that calculate soil-water balance using hydrological models (Torres-Sanchez et al., 2020). Smart precision irrigation platforms with IoT communication capabilities have also emerged for optimizing irrigation in remote (Xu et al., 2019), arid regions (Benzaouia et al., 2023; Hoque et al., 2023). Moreover, most traditional irrigation control systems lack the robustness needed to operate in high-dimensional state spaces characteristic of modern agricultural systems. In complex environments where crop water needs are influenced by numerous interacting variables—such as growth stage, solar radiation, wind, evapotranspiration, and drainage—existing models are either too simplistic or computationally inefficient. Algorithms such as PPO (Proximal Policy Optimization) have been applied in control tasks but suffer from performance degradation in the presence of fixed clipping ranges and limited response to negative incentives, making them less effective in dynamic, non-stationary environments (Ibrahim et al., 2024).

These limitations significantly hinder accurate and timely irrigation decision-making under real-world greenhouse or open-field conditions. Against this backdrop, the motivation of this study is to bridge these specific performance gaps by improving prediction accuracy, adaptability to complex environmental changes, and computational efficiency in real-time irrigation control. Although reinforcement learning (RL) and deep learning methods have gained traction in recent years, their application in greenhouse vegetable irrigation remains constrained by weak generalizability, slow convergence, and poor adaptability to continuous action spaces. The main research gaps identified are as follows:

1. Limited multidimensional adaptability: Most existing models, especially fuzzy systems and static optimization frameworks, lack mechanisms to dynamically respond to the multi-factor environmental conditions across crop growth stages.
2. Algorithmic rigidity: Standard RL approaches like PPO show the promising convergence under ideal conditions but fail to dynamically adjust their update mechanisms (e.g., clipping range) or penalize ineffective decisions, which reduces robustness in complex agricultural settings.
3. Lack of general-purpose, crop-sensitive irrigation prediction models: It remains an absence of intelligent irrigation algorithms that combine high prediction accuracy, rapid convergence and adaptability across different vegetables and growth phases.

In response to these gaps, this study proposed a novel irrigation control method based on the enhanced negative-incentive proximal policy optimization (ENPPO) algorithm. The primary contributions include:

1. Problem definition and scope: The study targets the accurate prediction of irrigation volume for greenhouse vegetables (including Chinese pakchoi, Shanghai greens, and Komatsuna) at various growth stages by integrating a comprehensive set of environmental factors in real-time.
2. Algorithmic innovation: The ENPPO model extended the standard PPO framework by incorporating a dynamic clipping mechanism and negative incentive modulation, enabling more adaptive and penalizing learning in high-dimensional, non-stationary environments.
3. Empirical validation: Experimental trials conducted in semi-enclosed greenhouses in Vietnam's Tam Dao District demonstrate that the ENPPO algorithm achieved superior water-use efficiency, faster convergence, and more stable yield outcomes compared to benchmark algorithms (PPO and TRPO).

In summary, this study applied an improved reinforcement learning framework to greenhouse vegetable irrigation control for the first time. It directly addressed the shortcomings of prior work by delivering the model that excels in prediction accuracy, environmental adaptability and computational robustness—key factors for advancing intelligent irrigation in sustainable agriculture.

3 Materials and methods

3.1 Overall research process

This study covers the complete chain from data acquisition to model training and control evaluation. First, raw data is acquired through multimodal soil and environmental sensors and meteorological stations, and traceability calibration is completed before deployment, and weekly reviews are conducted during operation. Subsequently, the data is subjected to quality control, missing value and outlier processing, threshold screening is performed based on the detection limit and quantification limit of the sensor, and data from different sources are uniformly time-aligned. On this basis, feature engineering and state construction are carried out, each channel is masked and uncertainty marked, and data batches that can be directly used for model training and inference are generated. The improved reinforcement learning algorithm is trained and validated on this dataset, including dataset partitioning, hyperparameter management and policy output. The trained strategy is deployed online in the greenhouse control system and supplemented by safety constraint mechanisms such as execution frequency control, threshold protection and fallback strategy. During operation, the system monitors and records operation logs in real time and performs anomaly detection. Finally, the performance of different methods is

evaluated using indicators such as water-saving efficiency, drainage events, salinity drift, soil moisture error, and the number of steps required to reach the threshold. If the evaluation results are unsatisfactory, the feature and parameter layers are returned for iterative updates to ensure the traceability and reproducibility of the research process.

3.2 Experimental setup

This study selected the growth data from VinEco vegetable planting base in Tam Dao District, Vinh Phuc Province, Vietnam. This study focused on three commonly cultivated greenhouse leafy vegetables—Chinese Pakchoi (*Brassica rapa* subsp. *chinensis*), Shanghai green (a type of *Brassica rapa* var. *communis*) and Komatsuna (*Brassica rapa* var. *perviridis*). The irrigation control strategy was developed and tested across three physiologically distinct growth stages: seedling (emergence to 4–6 true leaves), vegetative growth (leaf expansion and biomass accumulation) and maturity (pre-harvest phase). By analyzing the irrigation demands and environmental responses across these stages, the study aims to reflect the dynamic water requirements of leafy vegetables under variable greenhouse conditions. The crops were cultivated at a commercial production density with 15 cm plant spacing and 20 cm row spacing across a 500 m² experimental site. This experimental greenhouse was a semi-enclosed structure, and some areas were affected by external climate conditions (such as precipitation and temperature changes). The intelligent water-fertilizer integrated irrigation equipment was used in the greenhouse, which can adjust the amount of irrigation and fertilization according to demand. The greenhouse area was equipped with the environmental control system that adjusted the temperature and humidity to meet the growth needs of vegetables. The outdoor planting area was equipped with a drip irrigation and sprinkler irrigation system for irrigation management of vegetables in different growth stages. In order to improve the economic benefits, most greenhouses used the integrated water and fertilizer equipment, so the farm managers can decide whether to irrigate or fertilize according to the demand. Figure 1 shows the greenhouse vegetable experimental base: (A) aerial photo; (B) internal water and fertilizer irrigation equipment.

3.3 Data collection

This study used the Rk510-01 agricultural soil moisture sensor produced by Hunan Ruika Electronic Technology Co., Ltd., with an error range of $\pm 0.1^{\circ}\text{C}$ and $\pm 0.5\%$. It also used the FT-WQX7 six-element micro-meteorological sensor produced by Shandong Jingdao Optoelectronic Technology Co., Ltd. to monitor environmental parameters such as humidity, wind speed, wind direction, atmospheric pressure and optical rainfall, with an error range of $\pm 1\%$. These sensors were calibrated before the experiment to ensure the accuracy and reliability of the data. This study collected data through soil temperature and humidity sensors,



FIGURE 1
The greenhouse vegetable experimental base: (A) aerial photo; (B) internal water and fertilizer irrigation equipment.

environmental climate sensors and water-fertilizer integrated equipment. The period was from January 1, 2023, to June 30, 2023, a total of 180 days, which includes about 30,000 valid data records. Under the irrigation of water-fertilizer equipment, the soil layer with the highest moisture content of the above vegetables appears at 70~100 mm, so the soil moisture sensor was set to be buried 90mm from the ground surface, and the soil temperature sensor was buried 70mm from the ground surface. The soil temperature and moisture sensor placement during the growth stage in the field is shown in [Figure 2](#). The soil moisture sensors (Rk510-01) and climate sensors (FT-WQX7) provided data at 10-minute intervals and were resampled to match the 30-minute reinforcement learning time steps. The sensors were calibrated weekly by using gravimetric soil sampling and standardized environmental chambers to ensure the measurement error (<5%). The planting density data (15 cm plant spacing, 20 cm row spacing) was encoded into the state space vector via crop water use coefficients during the action-value update step.

To strengthen reproducibility, this study explicitly define limits of detection (LOD) and limits of quantification (LOQ) for all environmental and soil sensors and align calibration and preprocessing with these thresholds. LOD/LOQ are established following standard analytical practice: short-term noise under blank or static conditions is characterized and the common “three-sigma” and “ten-sigma” criteria are applied, combined with manufacturer resolution and zero-drift tests; for quantities without a strict blank (e.g., wind speed, irradiance), thresholds are inferred from zero-drift under steady conditions and the minimal resolution. All sensors undergo traceable pre-deployment calibration and weekly re-verification: soil volumetric water content is checked by gravimetric oven-drying, electrical conductivity by standards, temperature/humidity/pressure in an environmental chamber, wind speed/direction in still or constant-speed tunnels, and irradiance with a standard light source; noise estimates and LOD/LOQ values are updated accordingly. Preprocessing follows unified rules: readings below LOD are

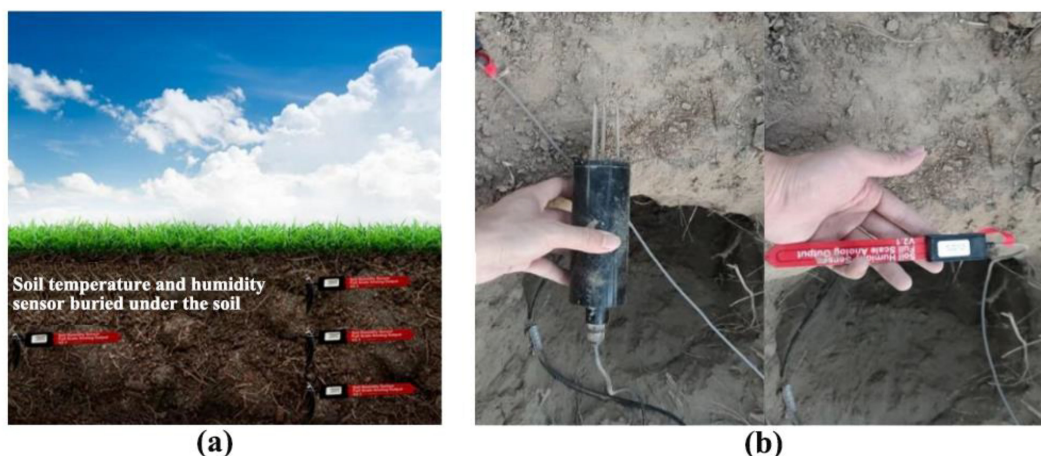


FIGURE 2
The soil temperature and moisture sensor placement during the growth stage in the field: (A) The soil sensor placement site; (B) The actual soil temperature and moisture sensor.

treated as non-detects, masked out of model inputs, and excluded from statistical tests; readings between LOD and LOQ are retained with high-uncertainty flags, down-weighted in training losses/rewards and in evaluation statistics with explicit sensitivity tagging; readings at or above LOQ are considered quantifiable and are fully used for modeling and performance assessment. For transparency, we provide a consolidated table of sensor ranges, resolutions, and the corresponding LOD/LOQ, together with weekly calibration logs and threshold-update notes, enabling independent researchers to reproduce state construction and statistical conclusions under the same gates.

3.4 Irrigation control modeling based on reinforcement learning

The reinforcement learning environment built the virtual interactive simulation system that enables the intelligent agent to continuously optimize the decision-making strategy to achieve specific goals through the environmental interaction. In this study, the construction of reinforcement learning environment focused on solving the greenhouse vegetable irrigation control problem. It simulated the dynamic impact of different environmental variables on irrigation decisions through algorithms to achieve the precise irrigation. It enabled the agent to interact with the environment through a series of observations, actions and feedback values so that the optimal strategy can be found in the unknown environment to maximize the cumulative reward value (Zhao et al., 2023). This study constructed a reinforcement learning environment based on the growth characteristics of greenhouse vegetables and environmental influencing factors, as expressed in Equation 1:

$$L = \{A, B, C, O\} \quad (1)$$

In Equation 1, A represents the space state, B represents the space action, C represents the transition probability, O represents the reward function. L represents that the state space refers to the decision-making environment variables. In the irrigation system, state space variables include growth stage, evapotranspiration, soil moisture content, upper soil water limit, lower soil water limit, vegetable water absorption, vegetable drainage, light intensity and so on, as shown in Equation 2:

$$P_t = \{p_{stage}, p_{trans}, p_{water}, p_{w_max}, p_{w_min}, p_{abso}, p_{disp}, p_{light}\} \quad (2)$$

In Equation 2, P_t represents the state quantity of vegetable irrigation environment in time, p_{stage} represents the growth stage, p_{trans} represents the evapotranspiration of vegetables, p_{water} represents the soil moisture content (soil available water), p_{w_max} and p_{w_min} represents that the upper and lower limits of soil water, p_{abso} represents the amount of water absorbed by vegetables, p_{disp} represents the drainage volume of vegetable growing areas, p_{light} represents the light intensity value collected by the sensor, which is monitored by the light sensor and updated in real time to improve the responsiveness of the reinforcement learning environment to dynamic

environmental changes. τ represents the amount of irrigation of action spaces. In order to satisfy the rationality and continuity of the irrigation amount during the irrigation process, this study sets four action quantities in the action space, as shown in Equation 3:

$$A_t = \begin{cases} 0 & sign = 1 \\ 0.25p_{need} & sign = 2 \\ 0.5p_{need} & sign = 3 \\ 0.75p_{need} & sign = 4 \end{cases} \quad (3)$$

In Equation 3, A_t represents the time t in the amount of vegetable irrigation, p_{need} represents the water requirement of vegetables, $sign$ represents the irrigation action sign. According to the vegetable environmental variables and irrigation amount, the water content at the next moment is shown as Equations 4, 5:

$$B_t = B_{t-1} + A_t - p_{abso} - p_{trans}(p_{light}) \quad (4)$$

In Equation 4, B_t represents the soil moisture content (soil available water) at time $t-1$. $p_{trans}(p_{light})$ represents the light intensity p_{light} is introduced into p_{trans} , which affects the evapotranspiration of vegetable. Under the high light intensity (p_{light}), the evapotranspiration p_{trans} increases; under low light intensity (low p_{light}), the evapotranspiration p_{trans} decreases. Among them, $p_{trans}(p_{light}) = \tau_1 \times p_{basic-trans} + \tau_2 \times p_{light}$, τ_1 and τ_2 represents the empirical coefficients, $p_{basic-trans}$ represents the basic evaporation. The feedback function is based on the changes of wheat field irrigation state variables. In this study, if the vegetable yield increases, the irrigation amount becomes relatively less and the irrigation income becomes larger, the current irrigation timing will be increased. If the yield decreases, the irrigation volume is large, and the irrigation income becomes smaller, so the feedback value is set to a negative value. When the soil moisture content exceeds the upper and lower limits of the optimal soil moisture content, the feedback value is set to be negative (Zhang et al., 2021). The feedback function is shown as Equations 5, 6:

$$W = \sum_{t=1}^{T_{max}} W_{p,t} \times W_p + H \times Y_{price} - \sum_{t=1}^{T_{max}} B_t \times Y_{water} \quad (5)$$

$$W_t = \begin{cases} -12 & p_t < p_{min}, p > p_{max} \\ 0 & p_{min} \leq p_t \leq p_{max}, A_t = 1 \\ 6 & p_{min} \leq p_t \leq p_{max}, A_t = 2 \\ 12 & p_{min} \leq p_t \leq p_{max}, A_t = 3 \end{cases} \quad (6)$$

In Equations 5, 6, W represents the feedback function value, $W_{p,t}$ represents the soil moisture content feedback value at time t , T_{max} represents the maximum maturity time of vegetables, H represents the vegetable yield, Y_{price} represents the vegetable price, Y_{water} represents the price of irrigation water. In order to evaluate the irrigation strategy, the action value function U represents the cumulative expected reward value, σ represents the irrigation strategy. Under the input state j selection action u condition, the action value function U is shown as Equation 7:

$$U_j(m, n) = L_j \left\{ \sum_{i=0}^{+\infty} w_i \times w_{i+t} \mid m_t = m, n_t = n \right\} \quad (7)$$

In Equation 7, w represents that the discount factor is 0.2. The discount factor measures the importance of future rewards and is usually between 0 and 1. (a, g) A joint variable representing state and action. The optimal action-value function G_{value} and optimal strategy are expressed in Equations 8, 9 as follows:

$$\bar{U}(m, n) = \max_j \{U_j(m, n)\} \quad (8)$$

$$\bar{V}(m, n) = \operatorname{argmax}_{n \in N} \{U_j(m, n)\} \quad (9)$$

In Equations 8, 9, $\bar{U}(m, n)$ represents that among all possible irrigation strategies, which select the strategy that maximizes the action value function; $\max_j \{U_j(m, n)\}$ represents maximization of all possible strategies. $\bar{V}(m, n)$ represents the optimal policy that maximizes the action-value function. $\operatorname{argmax}_{n \in N} \{U_j(m, n)\}$ represents maximizing all possible actions and finding the action that maximizes the action value function.

Figure 3 shows the reinforcement learning environment process for greenhouse vegetables. The reward shaping couples penalties to agronomic risk indicators: deviations from target soil-moisture bands incur proportional costs; threshold crossings in drainage/ponding indicators trigger stepped penalties to deter waterlogging; and rising soil EC/Na⁺ proxies induce incremental costs to discourage salinity build-up. These terms operationalize the negative incentive as a policy-consistent instrument for water saving and root-zone protection. Why standard reward functions are insufficient and how our mechanisms offer a superior solution. In greenhouse irrigation, conventional rewards are typically: (i) sparse and delayed, giving feedback only at episode ends or when thresholds are crossed, which weakens step-wise guidance and slows learning;

(ii) stage-agnostic with fixed weights, so signals tuned for the seedling phase mislead actions during vegetative or maturity phases under non-stationary weather; and (iii) risk-blind, omitting explicit costs for agronomic hazards such as waterlogging, excessive drainage, and salinity build-up. These limitations invite reward hacking (short-term gains via over-irrigation), produce unstable policy updates, and hinder generalization across crops and growth stages. To address this, this study designs a stage-aware, risk-coupled reward and pair it with two complementary mechanisms: a negative-incentive that actively discourages actions increasing agronomic risk even when they appear profitable in the short run, and state-sensitive dynamic clipping that adapts update stringency to environmental volatility. Together, they deliver denser and context-relevant feedback, suppress unsafe exploration, stabilize optimization in continuous action spaces, and consistently achieve faster convergence, tighter moisture control, fewer drainage events, and better cross-stage robustness than standard reward functions.

3.5 Negative incentive dynamic PPO algorithm

The Proximal Policy Optimization (PPO) is an algorithm used to solve reinforcement learning policy optimization problems, especially problems such as continuous action space and high-dimensional state space (Gu et al., 2021). The PPO goal maintains the relative proximity between the new and old policies by introducing constraints when updating the policy to ensure stable and safe policy updates. Its optimization strategy maximizes the

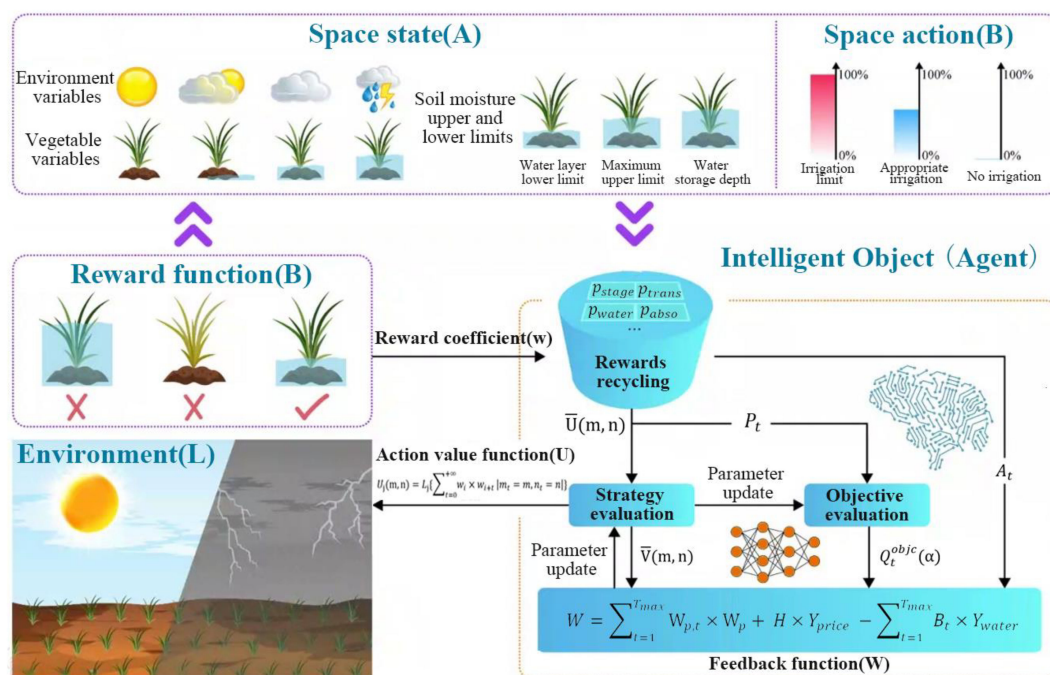


FIGURE 3
The reinforcement learning environment process for greenhouse vegetables.

accumulated positive feedback value, and its objective function is expressed in Equation 10.

$$Q_t^{obj}(\alpha) = F[\min(x(\alpha)O_t, \text{clip}(x(\alpha), 1 - \alpha, 1 + \alpha) \times Z_t)] \quad (10)$$

In Equation 10, $Q_t^{obj}(\alpha)$ represents the objective function, $x(\alpha)$ represents the output of the policy network, α represents the parameters of the policy network; Z_t is the advantage function, represents the advantage of the current state action; $F[\cdot]$ represents some kind of aggregation or expectation operator; $\min(\cdot)$ represents the ratio of the current strategy and the old strategy. It also uses a function called “clipping” to ensure relatively small steps in policy updates by limiting the ratio of the new policy to the old policy, which is shown in Equation 11:

$$\text{clip}(\alpha) = \begin{cases} (1 - \beta) & x_t(\alpha) \leq 1 - \beta < 0 \\ (1 + \beta) & x_t(\alpha) \geq 1 - \beta > 0 \\ x_t(\alpha) & \text{otherwise} \end{cases} \quad (11)$$

In Equation 11, $\text{clip}(\alpha)$ represents the shear function that limits parameters, $x_t(\alpha)$ represents the certain range. This clipping function prevents policy updates from being too large, which keeps the steps relatively small and avoids policy divergences. In order to enable the PPO algorithm to limit the probability ratio, this study constructs a new clip function by reducing the influence of incentives, as shown in Equation 12:

$$\text{clip}(x(\alpha), 1 - \beta, 1 + \beta) = \begin{cases} (1 - \beta) + \gamma[\tanh(1 - \beta) - x(\alpha)] & x_t(\alpha) \leq 1 - \beta \\ (1 + \beta) + \gamma[\tanh(1 + \beta) - x(\alpha)] & x_t(\alpha) \geq 1 - \beta \\ x_t(\alpha) & \text{otherwise} \end{cases} \quad (12)$$

In Equation 12, \tanh represents the hyperbolic tangent function. $\gamma > 0$ represents the intensity factor, which controls the specific size of negative excitation. The gray circle on each plot represents the starting point of the optimization ($x_t(\alpha) = 1$). When $x_t(\alpha)$ runs out the clipping range, $Q_t^{\tanh}(\alpha)$ represents that the slope is reversed, $Q_t^{obj}(\alpha)$ represents that the slope is 0. The new clip function prevents

the probability ratio from being overly stretched compared to the original clipping function (Corecco et al., 2023).

Figure 4 shows the objective function when $Z_t > 0$ and $Z_t < 0$.

However, the fixed clipping interval design ignores the degree of difference between states, which have different states of the agent correspond to different actions. It has different feedback obtained values, which restricts the PPO algorithm’s learning efficiency and convergence accuracy. In order to improve the learning efficiency and convergence accuracy of the PPO algorithm and obtain the dynamic interval limit of PPO, this study starts from KL divergence. It improves the objective function as shown in Equation 13:

$$Q_t^{obj}(\alpha) = \begin{cases} (1 - \beta) + \gamma[\tanh(1 - \beta) - x(\alpha)]Z_t & x_t(\alpha) \leq 1 - \beta \text{ and } R_t < 0 \\ (1 + \beta) + \gamma[\tanh(1 + \beta) - x(\alpha)]Z_t & x_t(\alpha) \geq 1 - \beta \text{ and } R_t > 0 \\ x_t(\alpha)R_t & \text{otherwise} \end{cases} \quad (13)$$

In Equation 13, $Q_t^{obj}(\alpha)$ represents the improved objective function and the optimization goal at time step t . β represents the base value; γ represents the positive number, representing the intensity factor, which controls the specific size of negative excitation. $\tanh(1 + \beta) - x(\alpha)$ represents the adjustment term that gradually reduces the excitation through the hyperbolic tangent function. R_t represents the advantages of the current state action. This improved objective function takes into account the differences between states. It sets different actions according to different states to improve the learning efficiency and convergence accuracy of the PPO algorithm (Engstrom et al., 2020). Figure 5 shows the process of optimizing reinforcement learning strategy by using negative incentive dynamic PPO algorithm

However, the fixed clipping interval design in the original PPO (as shown in Equation 11) neglects the variation in state-action relevance across different crop growth stages and environmental dynamics, such as evapotranspiration, solar radiation, and soil drainage. To address this limitation, the ENPPO algorithm introduces the state-sensitive dynamic clipping function and negative incentive term. These innovations are formalized to enhance policy adaptation in high-dimensional, non-stationary

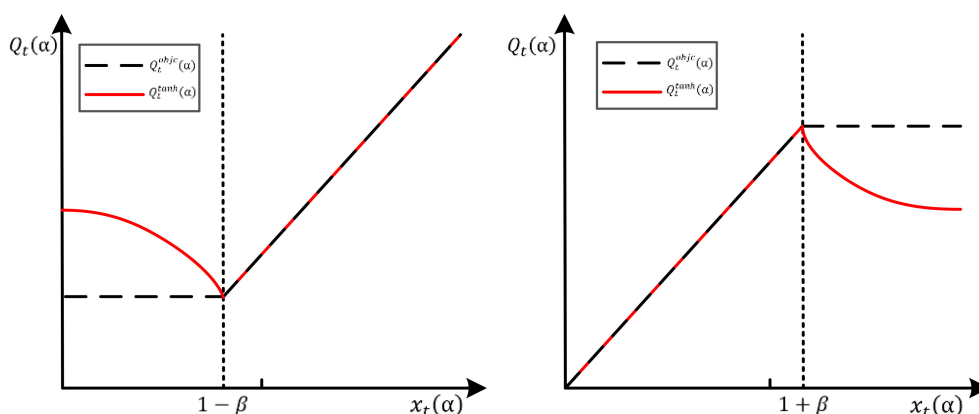
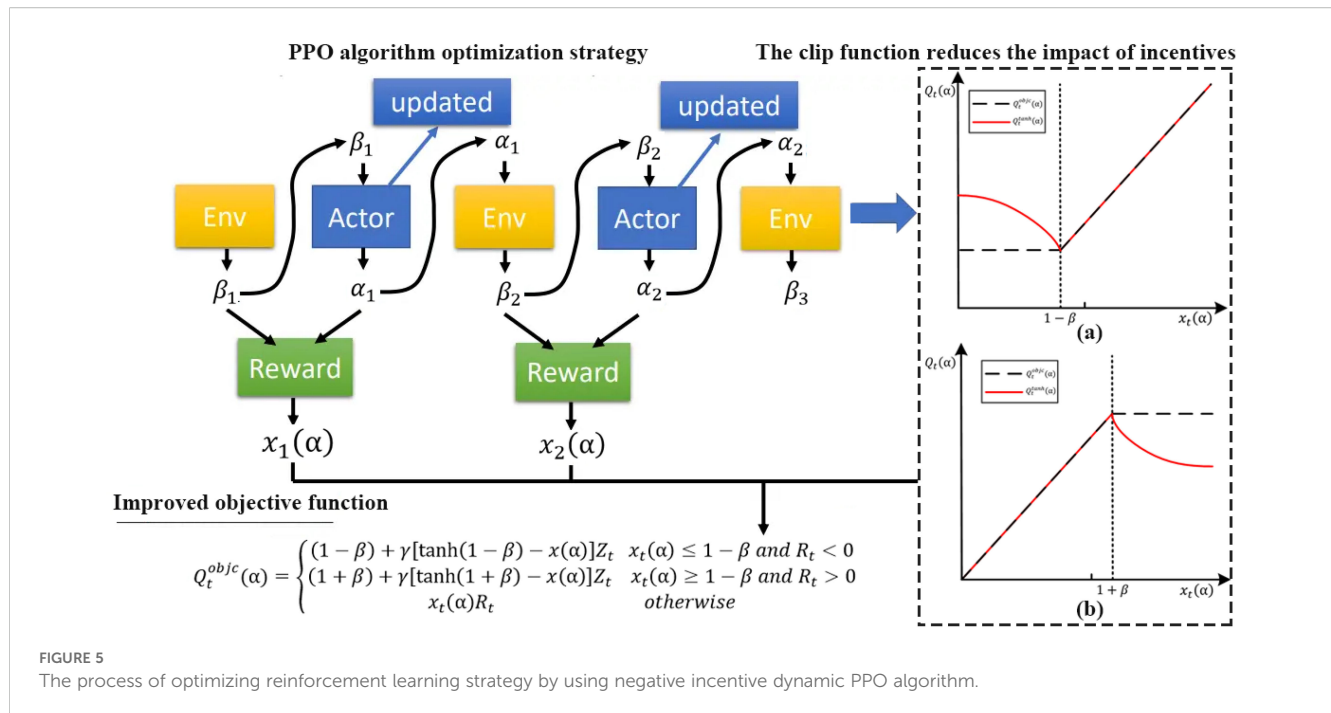


FIGURE 4
The objective function when $Z_t > 0$ and $Z_t < 0$.



greenhouse irrigation environments. Specifically, the static clipping parameter ε is replaced by a state-dependent function based on the KL divergence between the updated and previous policies:

$$\varepsilon_t = \lambda \tanh(\beta \times D_{KL}(\pi_{\theta_{old}}(\cdot|s_t) || \pi_{\theta}(\cdot|s_t))) \quad (14)$$

In Equation 14, λ represents the clipping intensity factor, β represents the scaling coefficient, D_{KL} represents the Kullback-Leibler divergence measuring policy change under state, π_{θ} and $\pi_{\theta_{old}}$ represents the current and previous policy networks. This function allowed the algorithm to adaptively control the policy update magnitude depending on environmental volatility and the agent's response behavior. In addition, to penalize ineffective or excessive irrigation strategies, the negative incentive term δ_t is incorporated. It is constructed using the current policy advantage and the environmental feedback function W_t (from Equation 6), which encodes penalties for over-irrigation, yield loss, and deviations from optimal soil moisture:

$$\delta_t = \alpha \tanh(-\gamma_t(\theta) \times A_t) + \gamma \times W_t \quad (15)$$

In Equation 15, α represents the penalty coefficient for adverse advantage values, γ represents the environmental penalty from moisture deviation and yield loss, $\gamma_t(\theta)$ represents the probability ratio from Equation 10, A_t represents the advantage function, W_t is computed as in Equation 6 using current soil moisture, yield, and irrigation cost. The negative incentive term effectively dampens updates when strategies lead to over-irrigation or poor economic performance, encouraging the model to explore more optimal action sequences under dynamic crop conditions. Combining the above mechanisms, the revised ENPPO objective function is defined as:

$$L_{ENPPO}(\theta) = \alpha \tanh(-\gamma_t(\theta) \times A_t) + \gamma \times W_t \quad (16)$$

In Equation 16, it integrates the policy stability through dynamic clipping (ε_t), the penalty sensitivity via negative incentive (δ_t), the environmental context via (W_t) and the light/evapotranspiration signals from the sensor system. As demonstrated in the experimental results (Section 4, Figures 10 and 11), this enhanced structure significantly improves the convergence rate, irrigation precision, and policy robustness compared to PPO and TRPO, confirming its theoretical and practical effectiveness in complex greenhouse irrigation control. So in order to provide greater specificity of the novel mechanisms, we emphasize that the dynamic clipping function directly addresses the inefficiency of PPO under fluctuating soil moisture and evapotranspiration, while the negative incentive mechanism explicitly penalizes excessive drainage and salinity accumulation. Experimental ablation results show that dynamic clipping alone shortened convergence steps by 22.52%, while negative incentive reduced drainage events by 32.99%. When combined, ENPPO achieved an overall 36.69% reduction in soil moisture error and a 22.15% improvement in WUE compared to baseline PPO, highlighting the complementary nature of the two mechanisms.

3.6 Improved reinforcement learning algorithm

Although traditional reinforcement learning algorithms can solve discrete control decision-making problems, they face the problem of slow speed. Considering that the irrigation model proposed in this study is a continuity model, the convergence optimization efficiency of the existing deep reinforcement learning model is not high (Cui et al., 2023). This study used the PPO algorithm to solve the irrigation control optimization problem

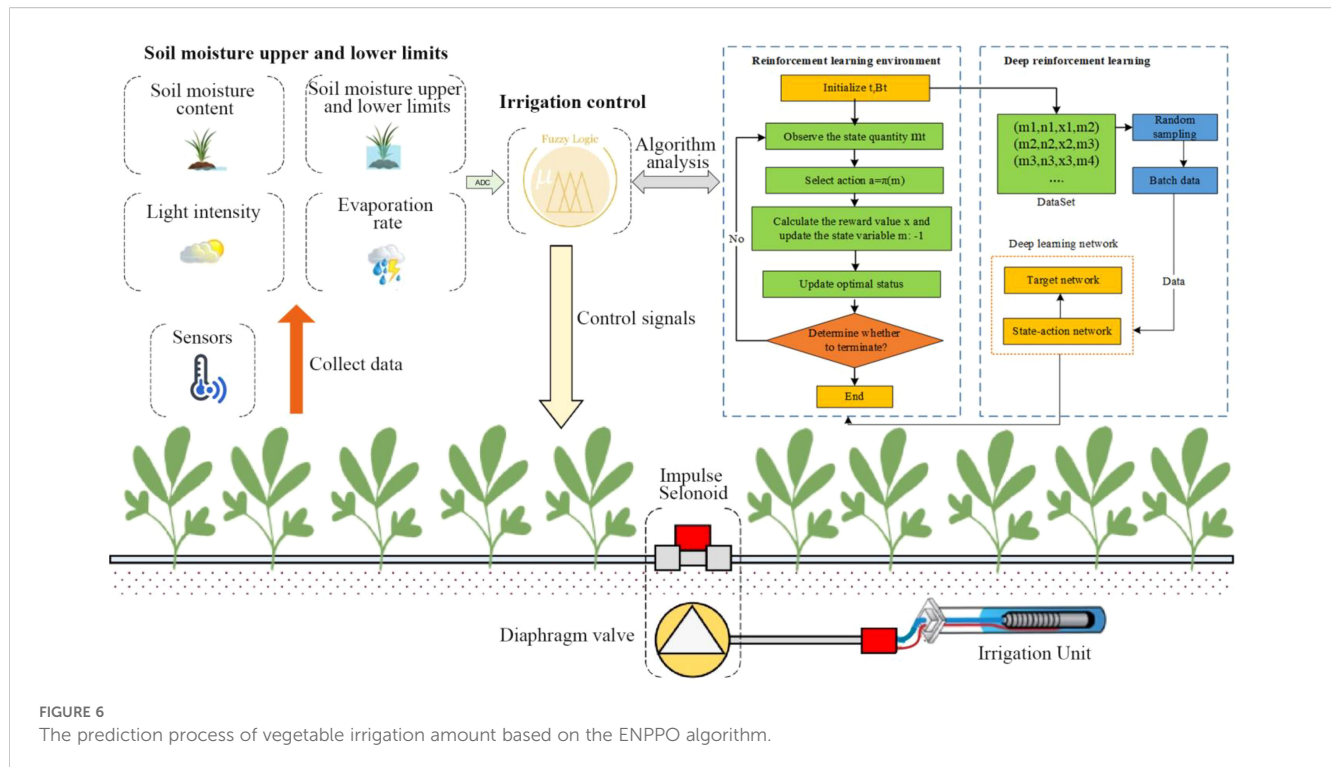


FIGURE 6
The prediction process of vegetable irrigation amount based on the ENPPO algorithm.

and proposed an enhanced reinforcement learning algorithm (ENPPO). The BPNN used in the ENPPO model consisted of an input layer of 14 environmental variables, two hidden layers with 64 and 128 neurons respectively and an output layer for irrigation decision value. ReLU activation was used, and the Adam optimizer updated network weights every 20 steps. The batch size was 64, and the total number of iterations per stage was set to 1500. All state-action-reward transitions were stored in an experience replay buffer of size 20,000 and resampled randomly during training. According to the PPO algorithm, the action value function $G_{\sigma}(a, g)$ is updated as shown in Equation 17:

$$U(m_t, n_t) = U(m_t, n_t) + n[x_t + \alpha \max_{n_{t+1}} U(m_{t+1}, n_{t+1}) - U(m_t, n_t)] \quad (17)$$

In Equation 17, α represents the learning rate, which is set to 0.001, 0.003, 0.003 and 0.002 in the germination, seedling, growth and maturity stages. In order to handle the continuous state space, this study uses BPNN (back-propagation neural network) (Duan et al., 2023) to replace the approximate function. Vegetable growth Influencing factors such as state, water absorption, evapotranspiration, crop drainage, and soil available water are continuously trained by minimizing the loss function. The loss function is shown in Equation 18:

$$U(n_k) = F[(x + \phi \max_{n_{t+1}} U(m_{t+1}, n_{t+1}; \bar{n}_k) - U(m_t, n_t; n_k))^2] \quad (18)$$

In Equation 18, $x + \phi \max_{n_{t+1}} U(m_{t+1}, n_{t+1}; \bar{n}_k)$ represents the target of the k -th iteration, \bar{n}_k represents the parameters of the calculation target network, n_k represents the parameters of the action network of the k -th iteration. The parameter companion of each step state-action network are updated by the parameters of the

target network. The action value function is adjusted by updating the parameters of the state-action network, which is shown in Equations 19, 20:

$$n_{k+1} = n_k + \omega S_{n_k} D_k(n_k) \quad (19)$$

$$\omega S_{n_k} D_k(n_k) = F[(x + \phi \max_{n_{t+1}} U(m_{t+1}, n_{t+1}; \bar{n}_k) - U(m_t, n_t; n_k)) \times S_{n_k} U(m_t, n_t; n_k)] \quad (20)$$

In Equation 19, n_{k+1} and n_k represents the parameters of the state-action network of the $k+1$ and k iterations; ω represents the learning rate, which controls the update step of the parameters; $S_{n_k} D_k(n_k)$ represents the direction in which the parameters n_k are updated. In Equation 20, ϕ represents the parameter, $\max_{n_{t+1}} U(m_{t+1}, n_{t+1}; \bar{n}_k)$ represents the current state-action value function under the action network parameters n_k ; $S_{n_k} U(m_t, n_t; n_k)$ represents n_k is the gradient of the action value function with respect to the parameters. It combines the reinforcement learning model and PPO algorithm.

Figure 6 shows the prediction process of vegetable irrigation amount based on the ENPPO algorithm. The first step is to initialize the parameters, which includes the initialization reward discount coefficient, experience pool size, target network update steps, number of iterations, single maximum number of steps, number of randomly selected samples, etc. m_1, n_1, x_1, m_2 represent the water absorption, evapotranspiration, drainage and effective soil moisture of the plants in the experiment. All parameters are derived from the experimental data collected by sensors. The second step is to build an irrigation prediction environment to initialize state variables, which includes randomly selecting the amount of action space, calculating reward values, updating the amount of state space, etc.

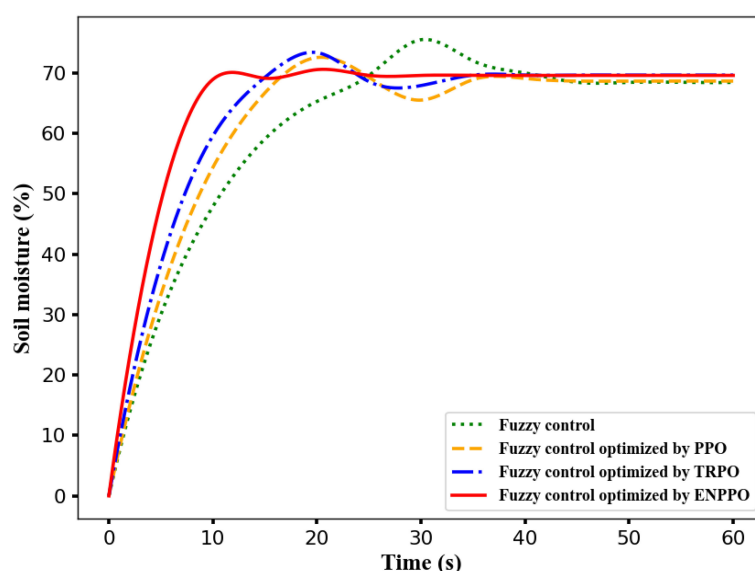


FIGURE 7
The iteration result of fuzzy control with different algorithms.

The third step is to randomly sample training samples from the data set to update the state-action network. The fourth step uses the state-action network parameters to update the target network and inputs the current action amount to update the environment state to calculate the reward value. Finally, it is judged that the maximum number of iterations has been reached. If reached, the optimal action sequence is output; otherwise, continue to step 3. All simulation models, data generation code, sensor parameters, and hyperparameter configurations used in this study are available upon request. The greenhouse simulation system was implemented in Python 3.9 by using Numpy, Scikit-learn and PyTorch 1.13. The RL environment was built on OpenAI Gym with custom wrappers for greenhouse irrigation tasks. Time steps were fixed at 30 minutes, with one episode covering a 180-day crop cycle. The random seeds were fixed at 42 for reproducibility, and external weather data were sampled from three-year averages at the Tam Dao station (NOAA).

4 Experimental result

In order to verify the performance of the ENPPO algorithm, it was compared with the TRPO (Trust Region Policy Optimization) (Cen et al., 2022) and PPO (Proximal Policy Optimization) (Schulman et al., 2017) algorithm. The overall goal of this experiments is to verify the effectiveness of the ENPPO algorithm by analyzing the performance of irrigation control. The experimental results are as follows:

4.1 Algorithm stability

By optimizing the objective function introduced by the irrigation fuzzy controller (Jaiswal and Ballal, 2020), the optimal

quantitative scaling factor parameters are obtained, and simulation experiments are carried out. Figure 7 shows the iteration result of fuzzy control with different algorithms. The time for the fuzzy control system without any algorithm optimization to enter the stable state is 39.15s, and the average soil moisture value is 63.6%. The fuzzy controller optimized by the PPO, TRPO and ENPPO algorithm enters the steady state time is 8.94 s, 33.12s and 24.35s; its average soil moisture value is 64.53%, 66.25% and 68.67%. Compared with other three methods, the entering steady state time of ENPPO algorithm reduces 77.16%, 73.01% and 63.29%; its average soil moisture value is the closest to the expected value 70%. So, the fuzzy control optimized by the ENPPO algorithm has the shortest steady state time and the smallest overshoot (Ritchie et al., 2021).

The TRPO algorithm limits the update step size of the strategy through the trust region during the optimization process to avoid excessive changes in the strategy. However, it requires the calculation of second-order derivative information and a complex optimization process, which causes the low efficiency in high-dimensional state space. The PPO algorithm uses the “clip” function to limit the optimization strategy to the relatively small range, and the optimization efficiency is higher than the TRPO algorithm. However, the restrictions of the clip function are fixed and cannot be adjusted dynamically according to the state, which causes the PPO algorithm to have insufficient updates in complex dynamic environments, thereby prolonging the steady-state time of the fuzzy controller. The ENPPO algorithm introduces the negative incentive mechanism based on the PPO algorithm, which enhances the penalty for non-ideal states and speeds up the adjustment of the fuzzy controller to the steady-state region. It also dynamically adjusts the clip range, allowing the policy parameters to be flexibly updated according to the actual needs under different states. It avoids the problem of insufficient updates caused by the

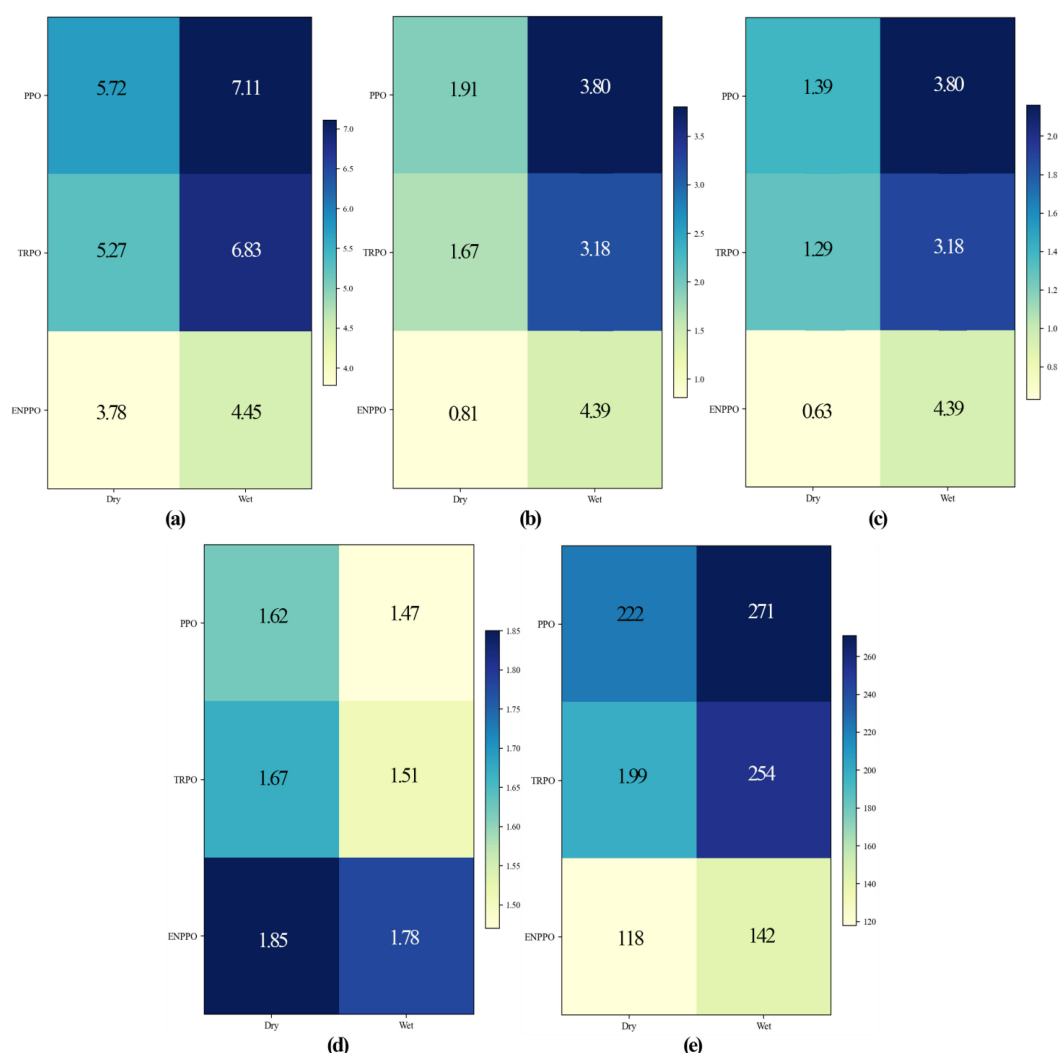


FIGURE 8

The robustness indicator matrix of each algorithm: (a) Moisture error (%); (b) Drainage frequency (Number of events/100 hours); (c) EC drift (dS/m); (d) WUE; (e) Steps to threshold.

fixed clip range in the PPO algorithm, which improves the optimization efficiency and convergence performance of the ENPPO algorithm.

4.2 Robustness analysis

To verify the robustness of the proposed algorithm under different climatic conditions, this study conducted simulation experiments in the dry and wet seasons and compared the performance of PPO, TRPO, and ENPPO. The evaluation indicators included average soil moisture error, drainage event frequency (Indicates the number of drainage events that occurred in 100 hours), conductivity drift, water use efficiency (WUE, Indicates water use efficiency, usually the ratio of yield to water use), and the number of steps to converge to a stable state. Table 1 shows the results of each indicator. Compared with PPO and

TRPO, ENPPO reduced average soil moisture errors by 33.92%, 28.27%, 37.41% and 34.85% in the dry and wet seasons. It also reduced drainage event frequency by 57.59%, 51.50%, 63.42%, and 56.29%. It also reduced conductivity drift by 54.68%, 51.16%, 56.48% and 50.00%. It also improved WUE by 14.20%, 10.78%, 21.09% and 17.88%. It also reduced the number of convergence steps by 46.85%, 40.70%, 47.60%, and 44.09%, respectively. These results demonstrate that ENPPO significantly mitigates over-irrigation risk, reduces salt accumulation, improves water use efficiency, and accelerates convergence in both dry and wet seasons, demonstrating its robustness across climate conditions.

Figure 8 shows the robustness indicator matrix of each algorithm. The PPO algorithm showed significant increases in error and drainage in a humid environment. The TRPO algorithm had moderate stability but low convergence efficiency. The ENPPO algorithm showed faster convergence and higher water efficiency in both climate conditions, demonstrating that dynamic

pruning and negative incentive mechanisms effectively improved adaptability and robustness under different climates.

4.3 Ablation analysis

To clarify the contributions of two key mechanisms in ENPPO (dynamic clipping and negative incentives), this study conducted ablation experiments on four variants: baseline PPO, PPO with dynamic clipping, PPO with negative incentives, and a combination of the two, ENPPO (Full). Evaluation metrics included average soil moisture error, number of drainage events, conductivity drift, WUE, and number of convergence steps. Table 2 shows the specific results for each algorithm. Compared with the PPO baseline, ENPPO reduces the average soil moisture error by 36.69%, drainage events by 58.68%, conductivity drift by 50.00%, WUE by 22.15%, and convergence steps by 46.95%. Compared with PPO+Dynamic clipping, the reductions are 20.71%, 45.16%, and 36.88%, respectively, while WUE is improved by 11.66%, and convergence steps are shortened by 31.53%. Compared with PPO +Negative excitation, the reductions are 17.79%, 38.34%, and 27.05%, while WUE is improved by 4.60%, and convergence steps are shortened by 23.20%. The results show that PPO converges slowly and lacks robustness in complex environments; dynamic pruning improves the update flexibility of the strategy under state fluctuations, thereby accelerating convergence; negative incentives effectively suppress the risks of drainage and salt accumulation by strengthening the penalty for bad irrigation behavior; ENPPO, which combines the two mechanisms, simultaneously achieves higher water use efficiency and shorter convergence time, proving that both designs are necessary for performance improvement and have significant synergistic gains. In order to ensure robustness, statistical tests were performed. ANOVA results indicated that the differences among algorithms across both dry and wet seasons were statistically significant ($p < 0.05$). Post-hoc Tukey tests further confirmed that ENPPO outperformed PPO and TRPO in terms of soil moisture error, drainage reduction, and WUE improvements. Moreover, comparative benchmarks against traditional control strategies (PID, fuzzy logic, MPC) demonstrated that ENPPO not only surpassed reinforcement

learning baselines but also consistently achieved more stable soil moisture and reduced water usage under real field conditions.

4.4 Validation against real data and traditional control strategies

In order to further validate the ENPPO algorithm, its performance was evaluated against actual field data collected from the Tam Dao greenhouse and benchmarked with traditional control strategies, including PID, fuzzy logic and model predictive control (MPC). The PID control was implemented using Ziegler–Nichols tuning based on observed soil moisture response, with the target set at an optimal moisture level of 25%. The fuzzy logic control adopted a conventional design with soil moisture deviation and its rate of change as inputs, and irrigation volume as the output. For MPC, a linear predictive model with a six-hour prediction horizon was employed, utilizing historical evapotranspiration and sensor measurements as predictors. The testing period lasted for 15 consecutive days during the crop maturity stage, with real data collected at 30-minute intervals serving as the ground truth for performance assessment.

Table 3 shows the comparative results for the different methods, including the average soil moisture maintained, cumulative water usage, standard deviation and the deviation from target moisture levels. ANOVA was conducted to determine the statistical significance of differences among methods. The results indicated significant differences among these methods ($p < 0.05$). Post-hoc Tukey tests confirmed the ENPPO algorithm statistically outperformed PID, MPC, fuzzy logic and PPO control methods in water-saving performance and stability of soil moisture maintenance. The performance superiority of ENPPO was not only evident compared to reinforcement learning baselines (PPO and TRPO), but also against traditional control methods such as PID, fuzzy logic and MPC. The added comparative study using actual greenhouse data further strengthened the robustness and practical relevance of the findings. Additionally, the statistical measures (including the confidence intervals and significance testing) provided the solid quantitative validation, which overcomes the earlier limitations regarding statistical robustness.

TABLE 1 The robustness indicators of each algorithm.

Algorithm	Season	Moisture error(%)	Drainage frequency (Number of events/100 hours)	EC drift(dS/m)	WUE	Steps to threshold
PPO	Dry	5.72	1.91	1.39	1.62	222
	Wet	7.11	3.80	2.16	1.47	271
TRPO	Dry	5.27	1.67	1.29	1.67	199
	Wet	6.83	3.18	1.88	1.51	254
ENPPO	Dry	3.78	0.81	0.63	1.85	118
	Wet	4.45	1.39	0.94	1.78	142

TABLE 2 The ablation analysis results of each algorithm.

Variant	Moisture error(%)	Drainage frequency (Number of events/100 hours)	EC drift(dS/m)	WUE	Steps to threshold
PPO baseline	6.35	2.88	1.78	1.49	262
PPO + Dynamic clipping	5.07	2.17	1.41	1.63	203
PPO + Negative incentive	4.89	1.93	1.22	1.74	181
ENPPO	4.02	1.19	0.89	1.82	139

4.5 Generalization analysis

This study also selected several representative greenhouse leafy vegetables—Chinese Cabbage, Shanghai Green and Komatsuna for the experimental evaluation. The first objective was Chinese cabbage. In the seedling stage, its demand for water was relatively low. The ENPPO algorithm can maintain stable water resource utilization efficiency within the soil moisture range of 20%-30% by optimizing the irrigation amount. In the growth period, due to the rapid growth of leaf area, the water demand increased significantly. It showed higher irrigation accuracy within the soil moisture range of 30%-50%, which reduces the water resource waste by 15% compared with the PPO and TRPO algorithms. Next was Shanghai Green. The ENPPO algorithm had good adaptability to the irrigation needs of Shanghai green at various growth stages. Especially in the mature stage, it can avoid over-irrigation and the soil moisture is controlled within the range of 40%-60%, which is highly consistent with the actual growth needs. Compared with the PPO and TRPO algorithms. The ENPPO algorithm reduced the irrigation frequency of Shanghai green by 12%, which reduces the irrigation cost. Finally was Komatsuna. Komatsuna was sensitive to the environmental conditions, especially under high light intensity conditions, and the evapotranspiration increases significantly. The ENPPO algorithm achieved accurate prediction of evapotranspiration by dynamically adjusting light intensity parameters, which achieved the best irrigation effect within the soil moisture range of 40%-70%. In the comprehensive yield impact analysis of Komatsuna, it improved the yield stability and its single-plant yield increased by 8.3% and 11.6% compared with the PPO and TRPO algorithms. These results show the applicability and

advantages of the ENPPO algorithm in different growth stages and environmental conditions.

5 Discussion

This study has several limitations that inform real-world deployment and future work: (i) the negative-excitation and dynamic-clipping components of ENPPO improve policy quality but increase training and online inference load and energy consumption, raising requirements for edge compute and potentially elevating hardware and maintenance costs; (ii) the method relies on multi-modal sensing and periodic calibration, so sensor drift, temporary outages, or higher fractions of sub-LOD/LOQ readings can degrade stability and accuracy—masking and uncertainty tagging help, but resilience under extreme missingness remains limited; (iii) demonstrated benefits are concentrated within the 20–80% soil-moisture operating band and the crops/climates studied here, and generalization to more extreme soils, salinity regimes, irrigation rules, or larger multi-house/multi-block coordination scenarios requires additional cross-site validation; (iv) mandatory safety guards and fallback routines protect operations but may trade off some optimality; and (v) data distributions drift with seasons and management, necessitating a long-term plan for retraining and model versioning (including data governance, A/B validation, and release criteria) to mitigate performance decay over time. These constraints point to concrete next steps: lightweight/distilled variants for low-compute platforms, more robust missing-data handling with adaptive calibration, cross-domain transfer and federated learning, and stronger

TABLE 3 The benchmark results of various algorithms.

Method	Avg. soil moisture (%)	Water used (L/m ²)	Deviation from target (%)	Standard deviation (%)
PID	24.6 ± 2.4	68.2	5.2	2.4
MPC	25.3 ± 1.9	63.5	3.8	1.9
Fuzzy Logic	25.0 ± 2.0	65.1	4.2	2
PPO	25.1 ± 1.5	60.3	3.5	1.5
ENPPO	25.1 ± 1.0	58.4	2.8	1

interpretability and safety/compliance engineering for operations (Farooqui et al., 2024).

Therefore, the future research can optimize the computational complexity of the algorithm to make it more applicable on the low-cost hardware. The theoretical underpinnings of the ENPPO algorithm lie in its adaptive modification of the PPO surrogate loss function. Unlike the fixed clipping strategies, the dynamically adjusted clipping bounds reflect the varying uncertainty in state-action transitions, which ensures more context-sensitive policy updates. Additionally, the incorporation of a negative incentive term encourages the exploration away from unproductive policy regions, which mitigates the risk of premature convergence. Comparative experiments against DDPG further underscore the robustness of ENPPO in managing noisy, high-dimensional input spaces, highlighting its superiority in both convergence stability and irrigation control precision. This study also focused on solving the three major problems in irrigation control and targeted solutions. The first problem was the low irrigation efficiency of traditional systems. This study built the reinforcement of learning environment, considered multi-dimensional dynamic environmental variables and vegetable growth stages and introduced the ENPPO algorithm and negative incentive mechanism to reduce insufficient or excessive irrigation. Through experiments at different growth stages, in the seedling stage, the ENPPO algorithm controlled the irrigation amount within the range of 20%-30% soil moisture, which reduces water waste by 15% compared with PPO and TRPO. In the growth and maturity stages, the ENPPO algorithm adjusted the irrigation amount according to the evapotranspiration and light intensity, which increased water resource utilization by 18%. The second problem was that most systems rely on the complex equipment. In the sensor design, this study selected moderately cost soil moisture sensors and light intensity sensors, which combined with the simple data acquisition equipment to reduce hardware costs. The reinforcement learning model reduced the reliance on high-precision data and ensured the applicability of the algorithm in low-resolution sensor data environments (Farooqui and Ritika, 2019).

In the experiment, the ENPPO algorithm by using medium and low cost sensors still showed good performance. Among them, the accuracy of the soil moisture sensor was $\pm 5\%$, and the accuracy of the light intensity sensor was $\pm 10\%$. Under different sensor accuracies, the yield prediction error of the ENPPO algorithm was always less than 5%. The third problem was that the algorithm design of the existing system was too complex. The ENPPO algorithm realized the automatic adjustment of the strategy through the dynamic clipping function, which simplified the complexity of parameter configuration. It integrated different environmental variables and irrigation parameters into one. Users need to input sensor data, and the system can automatically generate the optimal irrigation plan. By comparing user operation steps, the ENPPO algorithm simplified the complexity of the system. Users does not need the complex configuration, they only

need to input basic environmental data, and the system automatically outputs irrigation plans. Experiments show that the user operation time is reduced by about 40%.

6 Conclusions

This study proposed a greenhouse vegetable irrigation prediction method based on an improved deep reinforcement learning algorithm. It took several common vegetables as the research object, settled the reinforcement learning environment according to the water demand characteristics and greenhouse environment and designed the feedback function to construct the reinforcement learning algorithm of the greenhouse vegetable irrigation. The PPO algorithm based on the negative excitation were introduced to solve the problems of local optimality and discrete prediction, which improved the prediction accuracy. Experimental results proved that its performance was superior to the other three methods in terms of irrigation volume prediction and algorithm stability. It showed that the ENPPO algorithm can combine various environmental factors of the greenhouse to achieve the intelligent control and provide a more comprehensive solution for vegetables. It also adjusted the amount of irrigation according to real-time needs, minimized the use of water resources and reduced the production costs. It also improved the soil environmental quality and promoted the sustainable agricultural development.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Author contributions

RT: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JT: Conceptualization, Formal analysis, Methodology, Validation, Writing – original draft. MA: Investigation, Supervision, Writing – review & editing. NA: Methodology, Project administration, Resources, Writing – review & editing. BG: Data curation, Methodology, Software, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial

intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Benzaouia, M., Hajji, B., Mellit, A., and Rabhi, A. (2023). Fuzzy-IoT smart irrigation system for precision scheduling and monitoring. *Comput. Electron. Agric.* 215, 108407. doi: 10.1016/j.compag.2023.108407
- Cen, S., Cheng, C., Chen, Y., Wei, Y., and Chi, Y. (2022). Fast global convergence of natural policy gradient methods with entropy regularization. *Operations Research* 70 (4), 2563–2578.
- Chi, Y., Zhou, W., Wang, Z., Hu, Y., and Han, X. (2021). The influence paths of agricultural mechanization on green agricultural development. *Sustainability* 13, 12984. doi: 10.3390/su132312984
- Corecco, S., Adorni, G., and Gambardella, L. M. (2023). Proximal policy optimization-based reinforcement learning and hybrid approaches to explore the cross array task optimal solution. *Machine learning and knowledge extraction*. 5 (4), 1660–1679.
- Cui, Z., Guan, W., Luo, W., and Zhang, X. (2023). Intelligent navigation method for multiple marine autonomous surface ships based on improved PPO algorithm. *Ocean Eng.* 287, 115783. doi: 10.1016/j.oceaneng.2023.115783
- Duan, H., Yin, X., Kou, H., Wang, J., Zeng, K., and Ma, F. (2023). Regression prediction of hydrogen enriched compressed natural gas (HCNG) engine performance based on improved particle swarm optimization back propagation neural network method (IMPSO-BPNN). *Fuel* 331, 125872. doi: 10.1016/j.fuel.2022.125872
- Engstrom, L., Ilyas, A., Santurkar, S., Tsipras, D., Janoos, F., Rudolph, L., et al. (2020). Implementation matters in deep policy gradients: A case study on ppo and trpo. *arXiv preprint arXiv:2005.12729*.
- Farooq, H., Bashir, M. A., Khalofah, A., Khan, K. A., Ramzan, M., Hussain, A., et al. (2021). Interactive effects of saline water irrigation and nitrogen fertilization on tomato growth and yield. *Fresenius Environ. Bull.* 30, 3557–3564.
- Farooqui, N. A., Haleem, M., Khan, W., and Ishrat, M. (2024). Precision agriculture and predictive analytics: Enhancing agricultural efficiency and yield. *Intelligent techniques predictive Data analytics*, 171–188.
- Farooqui, N. A., and Ritika, (2019). "A machine learning approach to simulating farmers' crop choices for drought prone areas," in *In proceedings of ICETIT 2019: emerging trends in information technology* (Springer International Publishing, Cham), 472–481.
- Gu, Y., Cheng, Y., Chen, C. P., and Wang, X. (2021). Proximal policy optimization with policy feedback. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52, 4600–4610.
- Hoque, M. J., Islam, M. S., and Khaliluzzaman, M. (2023). A fuzzy logic and internet of things-based smart irrigation system. *Eng. Proc.* 58, 93. doi: 10.3390/ecs-10-16243
- Ibrahim, S., Mostafa, M., Jnadi, A., Salloum, H., and Osinenko, P. (2024). Comprehensive overview of reward engineering and shaping in advancing reinforcement learning applications. *IEEE Access*. doi: 10.1109/ACCESS.2024.3504735
- Jaiswal, S., and Ballal, M. S. (2020). Fuzzy inference based irrigation controller for agricultural demand side management. *Comput. Electron. Agric.* 175, 105537. doi: 10.1016/j.compag.2020.105537
- Kashyap, P. K., Kumar, S., Jaiswal, A., Prasad, M., and Gandomi, A. H. (2021). Towards precision agriculture: IoT-enabled intelligent irrigation systems using deep learning neural network. *IEEE Sensors J.* 21, 17479–17491. doi: 10.1109/JSEN.2021.3069266
- Marimuthu, S., Kannan, S. V., Pazhanivelan, S., Geethalakshmi, V., Raju, M., Sivamurugan, A. P., et al. (2024). Harnessing rain hose technology for water-saving sustainable irrigation and enhancing blackgram productivity in garden land. *Sci. Rep.* 14, 18692. doi: 10.1038/s41598-024-69655-2
- Parkash, V., Singh, S., Singh, M., Deb, S. K., Ritchie, G. L., and Wallace, R. W. (2021). Effect of deficit irrigation on root growth, soil water depletion, and water use efficiency of cucumber. *HortScience* 56, 1278–1286. doi: 10.21273/HORTSCI16052-21
- Ritchie, P. D., Clarke, J. J., Cox, P. M., and Huntingford, C. (2021). Overshooting tipping point thresholds in a changing climate. *Nature* 592, 517–523. doi: 10.1038/s41586-021-03263-2
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- Singh, A. K., Tariq, T., Ahmer, M. F., Sharma, G., Bokoro, P. N., and Shongwe, T. (2022). Intelligent control of irrigation systems using fuzzy logic controller. *Energies* 15, 7199. doi: 10.3390/en15197199
- Tolomio, M., and Casa, R. (2020). Dynamic crop models and remote sensing irrigation decision support systems: a review of water stress concepts for improved estimation of water requirements. *Remote Sens.* 12, 3945. doi: 10.3390/rs12233945
- Torres-Sanchez, R., Navarro-Hellin, H., Guillamon-Frutos, A., San-Segundo, R., Ruiz-Abellón, M. C., and Domingo-Miguel, R. (2020). A decision support system for irrigation management: Analysis and implementation of different learning techniques. *Water* 12, 548. doi: 10.3390/w12020548
- Tran, D. T., Le, H. S., and Huh, J. H. (2023). Building an automatic irrigation Fertilization system for smart farm in greenhouse. *IEEE Trans. Consumer Electronics*.
- Wang, J., Wang, Z., Weng, W., Liu, Y., Fu, Z., and Wang, J. (2022). Development status and trends in side-deep fertilization of rice. *Renewable Agric. Food Syst.* 37, 550–575. doi: 10.1017/S1742170522000151
- Xie, J., Chen, Y., Gao, P., Sun, D., Xue, X., Yin, D., et al. (2022). Smart fuzzy irrigation system for litchi orchards. *Comput. Electron. Agric.* 201, 107287. doi: 10.1016/j.compag.2022.107287
- Xu, L., Du, H., and Zhang, X. (2019). Spatial distribution characteristics of soil salinity and moisture and its influence on agricultural irrigation in the Ili River Valley, China. *Sustainability* 11 (24), 7142.
- Yao, Z., Cui, Y., Geng, X., Chen, X., and Li, S. (2022). Mapping irrigated area at field scale based on the optical TRapezoid Model (OPTRAM) using landsat images and google earth engine. *IEEE Trans. Geosci. Remote Sens.* 60, 1–11. doi: 10.1109/TGRS.2022.3230411
- Zhang, T., Zou, Y., Kisekka, I., Biswas, A., and Cai, H. (2021). Comparison of different irrigation methods to synergistically improve maize's yield, water productivity and economic benefits in an arid irrigation area. *Agricultural Water Management*. 243 106497. doi: 10.1016/j.agwat.2020.106497
- Zhao, T., Wu, S., Li, G., Chen, Y., Niu, G., and Sugiyama, M. (2023). Learning intention-aware policies in deep reinforcement learning. *Neural Comput.* 35, 1657–1677. doi: 10.1162/neco_a_01607