Computers and Electronics in Agriculture 237 (2025) 110554

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Adaptive fertilizer management for optimizing nitrogen use efficiency with constrained reinforcement learning

Hilmy Baja [a,*], Michiel G.J. Kallenberg [a], Herman N.C. Berghuijs [b,c], Ioannis N. Athanasiadis [a]

[a] Artificial Intelligence Group, Wageningen University & Research, The Netherlands
[b] Plant Production Systems, Wageningen University & Research, The Netherlands
[c] Wageningen Environmental Research, Wageningen University & Research, The Netherlands

ARTICLE INFO

Keywords:
Nitrogen use efficiency
Nitrogen surplus
Reinforcement learning
Crop management

ABSTRACT

Optimizing nitrogen use efficiency (NUE) in crop production is crucial for sustainable agriculture, balancing the need to maximize yield while minimizing environmental impacts such as nitrogen loss and soil nutrient depletion. Reinforcement learning (RL) emerges as a potent, data-driven approach for achieving optimal farm management decisions, particularly in the context of fertilization, thereby facilitating optimal NUE. Previous literature of RL in crop management have predominantly focused on optimizing yield, profit, or nitrogen loss reduction. However, optimizing NUE has been largely overlooked despite its significance in preventing soil nutrient mining. In this study, we develop an RL environment in various aspects to investigate the capability of RL to optimize NUE through crop growth model simulations. We develop an RL agent with a novel NUE reward function and incorporates action constrains. We compare its performance against baseline methods and other RL agents trained with reward functions from previous literature. Additionally, we evaluate the robustness of our RL agent across various soil conditions, including different initial nitrogen content and drought-(in)sensitive soils. We find that the RL agent trained with our novel reward function is close to the optimal policy, although generalization to different soil texture scenarios prove to be challenging to the RL agent. Further, we identify several open challenges for future work pertaining to RL in crop management.

1. Introduction and background

Feeding a growing global population under the pressing challenges of climate change remains a formidable task. Ensuring sustainable agriculture involves not only increasing crop yields but also mitigating negative environmental impacts arising from excessive resource use. Nitrogen (N) fertilizers, in particular, play a critical role in boosting agricultural production; however, excessive application can lead to detrimental outcomes such as pollution of water bodies, greenhouse gas emissions, and soil degradation (Cui et al., 2010; Tubiello et al., 2015). Conversely, insufficient N input can compromise crop productivity and contribute to soil fertility loss through "soil mining" (Van der Pol, 1992). Balancing high yields with minimal environmental harm is thus an essential goal for sustainable agriculture (Lipper et al., 2014).

Within this balance, nitrogen use efficiency (NUE) stands out as a critical agro-environmental indicator, describing the ratio of N in harvested grain to the total N input (Fageria and Baligar, 2005; Norton et al., 2015). Improving NUE reduces nitrogen surplus (N_{surp}) — the difference between applied N and N in harvested grain — and thus mitigates a range of environmental problems (Zhang et al., 2015; Klages et al., 2020). In recognition of these imperatives, the EU Nitrogen Expert Panel (EUNEP) has developed practical guidelines aimed at assessing and enhancing NUE (EU Nitrogen Expert Panel, 2015). Meanwhile, governmental bodies such as the European Union, the Food and Agriculture Organization (FAO), the USDA (U.S. Department of Agriculture), and directives such as EU Green Deal, have introduced stricter regulations to ensure responsible fertilizer use (European Union, 1991, 2016; FAO, 2019; Flach and Selten, 2021; Fetting, 2020), emphasizing the need for solutions that deliver both high productivity and ecological stewardship.

Despite heightened policy interest, current nitrogen management strategies are often based on farmers' generational experience, empirical good practice or reactive assessments of plant health, which can fail to capture the complexity of daily field dynamics (Abbas et al., 2021; Blackshaw et al., 2004; Altenbach et al., 2003). This gap has motivated a shift toward more flexible, data-driven methods for decision-making (Fountas et al., 2015). Among such approaches, *Reinforcement Learning* (RL) stands out due to its ability to sequentially adapt fertilization decisions in response to feedback and contextual

* Corresponding author.
E-mail address: hilmy.baja@wur.nl (H. Baja).

https://doi.org/10.1016/j.compag.2025.110554
Received 27 January 2025; Received in revised form 10 April 2025; Accepted 12 May 2025
Available online 4 June 2025

cues from in-field conditions (Gautron et al., 2022a). This adaptability is especially relevant when optimizing NUE, as the decision to apply fertilizer — and how much — can shift considerably across various soil, climate, or crop conditions. Hence, the core challenge is the conflict between achieving high agricultural productivity and mitigating the environmental damage caused by non-optimal nitrogen applications under challenging conditions. To overcome this, we develop a system where the RL agent is explicitly trained with an NUE reward function, and we conduct experiments to assess its ability to balance yield improvements with minimal environmental effects.

A growing body of RL literature demonstrated potential for improving yields and profitability in agricultural decision-making (Goldenits et al., 2024). However, most of these approaches do not explicitly target NUE or similar agro-environmental objectives, but instead emphasize yield maximization or profit (Overweg et al., 2021; Kallenberg et al., 2023; Gautron et al., 2022b; Wu et al., 2022; Madondo et al., 2023; Turchetta et al., 2022). This constitutes a critical gap, given that an RL agent's reward function predominantly dictates the type of policy it learns (Eschmann, 2021). Optimizing for yield and profit alone may unintentionally overlook the risks of excessive N losses and their long-term effects on soil mining. In contrast, reward functions grounded in indicators like NUE and N_{surp} may be better suited to encourage sustainable intensification. Consequently, in this paper we aim to design an RL agent that explicitly incorporates these agro-environmental metrics into the reward function in the form of the NUE indicator.

Moving from intended principles to realistic management requires accommodating farmers' practical constraints. For instance, most farmers prefer a limited number of fertilizer applications. Moreover, fertilizing after specific phenological stages often yields negligible benefits while elevating the risk of N runoff (Iizumi and Ramankutty, 2016). Hard-coding these agronomic constraints into an RL environment can reduce exploration and limit the agent's ability to learn *why* certain actions are suboptimal (Liu et al., 2021). Instead, incorporating constraints in the RL agent's learning process may lead to agents that better recognizes these constraints. In this work we utilize *LagrangianPPO* — an extension of Proximal Policy Optimization (PPO) (Schulman et al., 2017) — that balances NUE-centered objectives with realistic operational limits (Fisher, 1981; Ji et al., 2023). By doing so, we aim to yield fertilization policies that not only maximize our NUE-centric objective, but also align with agronomic realities and nitrogen policies.

1.1. Research questions

Motivated by the urgent need to reconcile productivity and sustainability, as well as growing policy focus on NUE-based metrics, this paper concentrates on the following questions:

1. **Environmental performance:** How effectively does an RL agent trained with a NUE-centric reward optimize important agronomic and environmental metrics relative to state of the art RL reward functions and baselines?
2. **Policy adaptability:** How well does an RL agent trained with a NUE-centric reward adapt its fertilization policies to varying soil scenarios (e.g., soil type, initial N levels), while maintaining good metric performance?

2. Materials and methods

2.1. Overview

In this study, we investigated the capability of RL to optimize the NUE metric. We conducted a representable *in-silico* case study on rain-fed winter wheat. The case study is situated in the Lelystad region of the Netherlands, where we employed the latest version of the WOFOST crop growth model to simulate nitrogen (N) dynamics and yield formation. Calibrated parameters were derived from field experiments reported in Groot and Verberne (1991), ensuring that the model closely reflects local soil conditions, cultivar traits, and climate characteristics.

To optimize nitrogen fertilization strategies, we formulated an RL environment we call *CropGym*, in which an RL agent interacts with the WOFOST model in discrete weekly time steps, as we know a priori that a good policy requires sparse interventions. Specifically, at each time step, the agent receives state information such as soil N availability, phenological stages, and current weather conditions. It then decides how much N fertilizer to apply (or whether to skip fertilization) as an *action*. Our approach contrasts with traditional, rule-based fertilization schedules by enabling adaptive decisions based on real-time simulations of crop status and environmental factors.

A key contribution of this work is a novel reward formulation centered on agro-environmental indicators, namely nitrogen use efficiency (NUE) and nitrogen surplus (N_{surp}). To handle real-world constraints — such as limiting the number of fertilization events and preventing fertilization at late crop stages — we employ a variant of Proximal Policy Optimization (PPO) with constraints through the Lagrangian method (Schulman et al., 2017; Ji et al., 2023). This *LagrangianPPO* approach dynamically balances constraint satisfaction with maximizing our NUE-N_{surp}-oriented reward function.

We trained each RL agent using a single, representative soil profile calibrated for the Netherlands. We evaluated several RL agents under various soil conditions to test the generality of our method:

1. low and high initial N in the soil,
2. fast and slow draining soil profiles.

Each training run comprised multiple simulation episodes ($3M$ episode steps), where an episode spanned a single growing season from sowing to harvest. We utilize random simulated weather in the training runs and evaluate our results with historical weather from years 1981 to 2021.

We compared our RL agent trained with the NUE reward function against:

1. a baseline rule-based policy reflecting standard farmer practice,
2. a baseline *optimal* policy based on the reward function,
3. an RL agent trained on relative-yield[1] rewards,
4. an RL agent trained on yield-N-loss rewards,
5. an RL agent trained on profit-oriented rewards.

Performance was assessed primarily in terms of yield, NUE, and N_{surp}, thereby addressing our two main research questions.

This section proceeds as follows: Section 2.2 outlines the underlying problem setup and elaborates the formal mathematical problem formulation, including state/action spaces and the integration of WOFOST into the RL loop. Section 2.3 details the RL algorithmic components and the Lagrangian constraint method. Section 2.4 then describes our simulation environment and reward function, while Section 2.5 covers training protocols, baselines, and evaluation metrics. Finally, Sections 3 and 4 present and interpret the results in light of our agronomic and environmental objectives.

2.2. Problem definition

In this section we formalize the problem of sequential decision making in crop management as a Markov Decision Process (MDP). An MDP can be described with the tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma \rangle$, where $\mathcal{S}$ is the state space and $\mathcal{A}$ is the action space. $\mathcal{T}$ and $\mathcal{R}$ are the environment's transition function $\mathcal{T}(s_{t+1}|s_t, a_t)$ and reward function $\mathcal{R}(s_t, a_t, s_{t+1})$, respectively. γ is the discount factor, with ranges $[0, 1]$, which determines

[1] Relative compared to a zero-fertilization policy, used by Overweg et al. (2021), Kallenberg et al. (2023).

how much future rewards are valued compared to immediate rewards. In crop management problems, as with many real-world environments, the agent is not privy of the complete environment state. In addition to the standard MDP elements, $\mathcal{O}$ is introduced as the space of possible observations $o \in \mathcal{O}$, which has an observation function of $O(o_t|s_t, a_t)$. For constrained optimization problems, the MDP can be described as a Constrained MDP (CMDP). CMDP introduces the MDP element $\mathcal{C}$, formally $C_i(s_t, a_t)$, which represents the penalty or "cost" incurred for taking a constrained action a_t in state s_t. i represents the constraint functions $C_1, C_2, \ldots, C_i$ implemented in the CMDP.

Overall, the MDP of this problem can be described with the tuple $\mathcal{M} = \langle S, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{O}, \mathcal{C}, \gamma \rangle$. The RL agent seeks to find a policy that maximize cumulative reward, represented by the objective function

$$f(\theta) = \max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t \mathcal{R}_t(s_t, a_t)\right] \tag{1}$$

where $\mathcal{R}_t(s_t, a_t)$ is the reward function. We set a fixed planting and harvesting date, yielding a fixed horizon T. Also, we set γ to 1, so our objective function is the undiscounted maximum expected cumulative reward in a trajectory.

Further, we constrain the actions of the agent to obtain actionable fertilization policies by using constraint functions $C_i(s_t, a_t)$. In our problem, we define two constraint functions:

1. constraint for the number of fertilization actions an agent can perform in one growing season, and
2. constraint for *when* an agent can perform fertilization actions

which functions we refer to as C_1 and C_2, respectively. The constraint for C_1 we define as

$$C_1 = max(0, \text{fert}_t - \text{fert}_{\text{max}}), \tag{2}$$

where fert_t is the cumulative fertilization events at time t and fert_{max} is the desired maximum cumulative fertilization events in the growing season, which we set to n=4, to follow the number of fertilization events farmers typically perform in a growing year (Yang et al., 2022). Next, we define C_2 as

$$C_2 = \mathbb{I}(DVS \leq DVS_{start} \vee DVS \geq DVS_{end}), \tag{3}$$

where $\mathbb{I}$ is an indicator function that returns 1 if the condition inside is satisfied, 0 otherwise. DVS is the development stage of the crop, DVS_{start} and DVS_{end} are scalars that describe the window in the development stage when the agent is allowed to fertilize, which we set to 0.01 and 1, respectively. The two scalars indicate the stage where the crop has emerged ($DVS > 0$) and when the crop has reached anthesis/flowering ($DVS = 1$).

For RL agents trained with other reward functions, we also implement constraints for ranges of NUE and N_{surp}, following the advice of Turchetta et al. (2022): directly constraining environmental indicators during the training of an agent. The constraints we define as follows:

$$C_3 = \mathbb{I}(0.5 \leq \text{NUE} \leq 0.9), \text{and} \tag{4}$$

$$C_4 = \mathbb{I}(0 \leq \text{N}_{surp} \leq 40). \tag{5}$$

Overall, the agent's behavior is constrained by the above set of constraints C_i, and we ensure that the expected cumulative costs do not exceed a threshold d_i:

$$g(\theta) = \mathbb{E}\left[\sum_{t=0}^{T} C_i(s_t, a_t)\right] \leq d_i \ \forall i, \tag{6}$$

where we set d_i to 0, so none of the constraints are violated by the agent.

In the context of our problem, the tuples could be mapped as follows: S is the whole range of crop, soil and environmental states simulated by our CGM, some which are hidden to the agent. Elements in $\mathcal{O}$ are a subset of S; crop, soil and environmental states that the agent can observe. $\mathcal{A}$ represents the levels of fertilization. $\mathcal{T}$ is a simulation step of the CGM. T is the simulation duration. $\mathcal{C}$ is the penalty awarded to the agent for violating certain constraints. And finally $\mathcal{R}$ is a compound feedback consisting an evaluation of how efficient and how much yield the agent obtained in the growing season. We explain $\mathcal{R}$ in-depth in Section 2.4.3.

2.3. RL agent

In this section we describe the algorithm and network of the RL agent we use in our experimental setting. The concept of *LagrangianPPO* was first introduced by Tessler et al. (2018), introducing the concept of policy constraints through Lagrangian multipliers. In this paper, we adapt the *LagrangianPPO* implementation of *Safety-Gymnasium* (Ji et al., 2023). Specifically, we adapt functions that update the Lagrange multipliers and loss calculations into the framework of Stable Baselines 3 (Raffin et al., 2019) — modifying the base PPO algorithm using the clipped surrogate function.

2.3.1. The Lagrangian method

Lagrangian methods are common for training RL agents with constraints (Ji et al., 2023). In general, an adaptive penalty coefficient λ_i, was implemented to enforce constraints. This λ is updated with a rule as follows:

$$\lambda_i \leftarrow \lambda_i + \alpha_{lag}(g(\theta)), \tag{7}$$

where α_{lag} is the learning rate for λ_i and $g(\theta)$ is the cost function from Eq. (6).

To ensure the penalties from violating a constraint is learned by our PPO agent, the Lagrangian constraint is included in the loss function, where it is a term added to the loss we aim to minimize. The Lagrangian constraint loss term is defined as follows:

$$\mathcal{L}_{lagrangian}(\lambda_i) = \sum_{i=0}^{n} \lambda_i(g(\theta)), \tag{8}$$

where n is the total number of constraint functions. Hence, the loss function of the PPO agent becomes:

$$\mathcal{L} = \mathcal{L}_{policy} + c_v\mathcal{L}_{value} + c_e\mathcal{L}_{entropy} + \mathcal{L}_{lagrangian}, \tag{9}$$

where $\mathcal{L}_{policy}$ and $\mathcal{L}_{value}$ are the loss of the surrogate function and value loss, respectively. Meanwhile, c_v and c_e are weight coefficients for the value loss and the entropy loss, and we set as 0.5 and 0.01, respectively. These weight coefficients scale the losses to change the influence of each loss term. The remaining terms in Eq. (9) are explained in more detail in the original PPO paper (Schulman et al., 2017). In Table B.7, we describe the PPO hyperparameters we used in our experiments.

2.3.2. Lagrangian ppo for crop management

In order to incorporate Lagrangian constraints into the PPO algorithm, we adapted its general architecture to allow for a Lagrangian constraint calculation. In this section, we describe how we adapt the *LagrangianPPO* implementation from *Safety-Gymnasium* into the Stable Baselines 3 framework. PPO is an actor-critic algorithm: The actor network is responsible for selecting actions based on its learned policy $\pi_\theta(a|s)$, and the critic network estimates the state-value function $V_\theta(s)$. θ is the parameter of the network. For Lagrangian constraints, the architecture was modified by adding an additional critic network: the *constraint critic*. The constraint critic estimates the constraint functions based on the current state of the environment $C_\theta(s)$. Similar to the value function $V_\theta(s)$, we also calculate the generalized advantage estimation (GAE) for the constraint:

$$A_t = \sum_{t=0}^{T} \lambda_{GAE}\delta_t^c, \tag{10}$$

Table 1
Crop, soil and weather features that the agent receives from WOFOST as its observation space.

Feature	Description	Units
DVS	Development stage of the crop	[–]
TAGP	Above ground dry weight biomass	[kg/ha]
LAI	Leaf Area Index	[–]
TRA	Transpiration rate from plant canopy	[cm/d]
RFTRA	Reduction factor for transpiration	[–]
WSO	Dry weight storage organ	[kg/ha]
NamountSO	N amount in storage organ	[kg/ha]
NuptakeTotal	Total plant N uptake	[kg/ha]
Week	Week since planting	[week]
Naction	Number of actions (> 0) taken since planting	[–]
NO3	Soil nitrate content (array)	[kg/ha]
NH4	Soil ammonium content (array)	[kg/ha]
WC	Water content in different soil (array)	[cm]
SM	Root zone soil moisture (array)	[–]
NLOSSCUM	Cumulative N loss	[kg/ha]
RNO3DEPOSTT	Total nitrate deposition in soil	[kg/ha]
RNH4DEPOSTT	Total ammonium deposition in soil	[kg/ha]
IRRAD	Solar Irradiance	[J/m^2/d]
TMIN	Minimum Temperature	[°C/d]
RAIN	Daily Rainfall	[cm/d]

where λ_{GAE} is the GAE hyperparameter that controls the trade-off between bias and variance. δ_t^c is the temporal difference (TD) residual of the constraint at timestep t, calculated as:

$$\delta_t^c = c_{t,\pi} + C(s_{t+1}) - C(s_t), \tag{11}$$

where $c_{t,\pi}$ are the constraint violations of the current policy π at time t, and $C(s_t)$ and $C(s_{t+1})$ are the constraint value estimations and constraint value estimation of the next step, respectively. To incorporate these constraint calculations, we modify the PPO roll out buffer where we add additional elements that relate to the timing and frequency of fertilization actions.

2.4. Simulating crop responses

In this section we describe in detail how we simulate crop responses and the interface we developed for the simulator. World Food Studies (WOFOST, Van Diepen et al., 1989; De Wit et al., 2019) is a robust crop growth model that has been thoroughly validated (Ceglar et al., 2019). It is a key component in the MARS crop yield forecasting system[2] (Van der Velde and Nisini, 2019) and the Global Yield Gap Atlas[3] (van Bussel et al., 2015). WOFOST has been recently expanded to include a dynamic soil N module, called SNOMIN (Soil Nitrogen module for Organic and MIneral Nitrogen), which enables more complex soil-crop N processes (Berghuijs et al., 2024). This expansion allows for better exploration of sustainable N fertilization polices. WOFOST SNOMIN (henceforth will be referred to as "WOFOST" in this paper) distinguishes two N types, NO_3^--N (nitrate) and NH_4^+-N (ammonium). Also, WOFOST distinguishes different soil layers for N and water dynamics.

2.4.1. RL environment

We utilize the Python version of WOFOST, implemented in the Python Crop Simulation Environment (PCSE, de Wit, 2023), to simulate the crop responses for our RL agent. We design an RL interface utilizing the Gymnasium API (Towers et al., 2024). Our RL interface, *CropGym*,[4] directly communicates with WOFOST and includes additional features compared to the previous version (Kallenberg et al., 2023).

The RL environment automatically calculates NUE, N_{surp}, and other important agro-environmental indicators after each episode, and stores the values in the environment's *info* variable from *Gymnasium's step()* function. We also provide a pipeline to add random weather, utilizing *LARS-WG8.0* (Semenov et al., 2002), for training an RL agent. The preprocessing steps are documented in our code repository. For the task of optimizing NUE, the initial amount of N in the soil is important, as it affects the required fertilization policy for efficient NUE. So, we include a method to randomize the initial soil, randomizing around a mean and standard deviation that can be specified by the user.

WOFOST has a plethora of crop states that the agent can use. We selected a subset of the features, based on consultation with domain experts, that the agent can observe to solve our current task. These features are shown in Table 1. As WOFOST has multiple soil layers, some features are represented as arrays. We process the output of WOFOST, so the RL agent receives the *sum* of the array for each *NO3* and *NH4*, and the *mean* of the array for each *WC* and *SM*.

In WOFOST, a user can define the concentration of N in rain water (*NO3ConcR* and *NH4ConcR*, in mgN/L). This relates to the annual N deposition (explained further in Section 2.4.2). *CropGym* automatically converts the obtained N deposition for a certain year into daily N deposition based on the planned simulation days, then subsequently updates the parameter file for the following episode. This ensures that the variables *RNO3DEPOSTT* and *RNH4DEPOSTT* outputs the correct deposition amounts for a certain simulation year.

The agent's action space consists of 9 levels of N fertilization: $\mathcal{A} = \{10n \mid n = 0, 1, \ldots, 8\}$ kg/ha. This follows a farmer's common fertilization amount, where 80 kg/ha is a typical upper-bound for a single fertilization event.

To simplify the flattened vector that the RL agent observes, we aggregated the time series data following Kallenberg et al. (2023): the sequence of weather with a length of 3×7 (i.e., daily rain, temperature and solar irradiance) was processed into an average pooling layer, resulting in a vector size of 3×1. The crop features that had a length of 17×7 were shrunk to 17×1 by taking its last entry.

2.4.2. Agro-environmental indicators

In this section, we elaborate the agro-environmental indicators that we use as our RL agent's reward function. As noted in the previous section, and also following Berghuijs et al. (2024), we adopt the NUE definition from the EUNEP framework (EU Nitrogen Expert Panel, 2015). The framework defines NUE as the ratio of N in the crop grains to the effective amount of N applied as fertilizer:

$$NUE = N_{out}/N_{in}, \tag{12}$$

where they define range of optimal N use to be between 0.5 kgN/kgN and 0.9 kgN/kgN. A value lower than 0.5 kgN/kgN is defined as

[2] https://joint-research-centre.ec.europa.eu/monitoring-agricultural-resources-mars_en
[3] https://www.yieldgap.org
[4] https://cropgym.ai

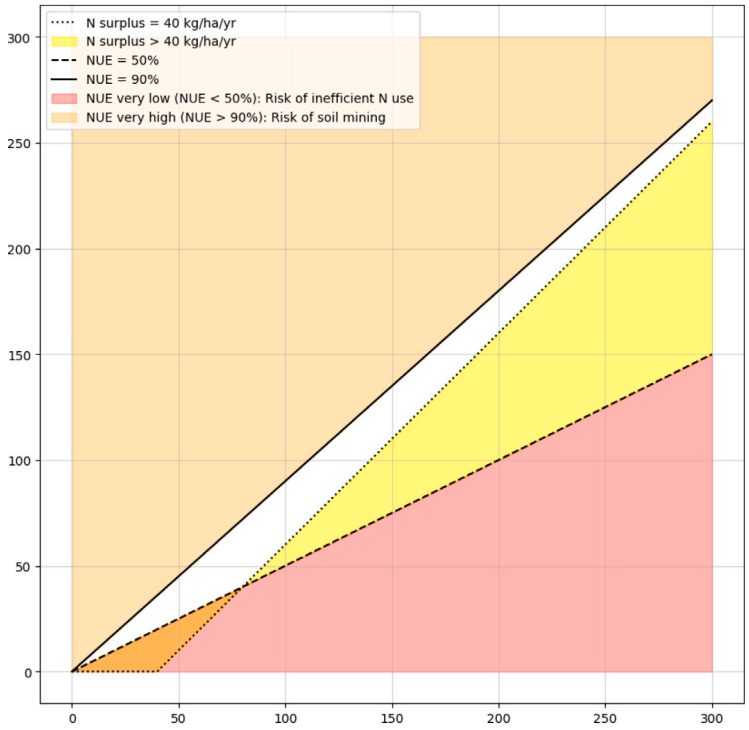

Fig. 1. The Nitrogen Use Efficiency framework plot used in this study.

inefficient N use, and a value higher than 0.9 kgN/kgN indicates a risk of soil mining. We calculate N_{out} in WOFOST as the simulated amount of N in the grains at harvest. In the remainder of the paper, we will show NUE as unitless for simplicity. N_{in} is calculated as follows:

$$N_{in} = N_{seed} + N_{depo} + N_{applied}, \tag{13}$$

where N_{seed} is the assumed amount of N in the sown seed of winter wheat Silva et al. (3.5 kgN/ha, 2021). N_{depo} is the annual daily atmospheric N deposition rate (in kgN/ha) calculated from the Dutch national deposition statistics (CLO, 2022) and daily precipitation rates of different years. For this paper, we calculate yearly N deposition from the trend as shown in Fig. C.6. $N_{applied}$ is the total amount of N fertilizer applied in the growing season in kgN/ha. In this study, we only utilize synthetic fertilizers that does not include organic matter.

Further, we calculate N_{surp} as follows:

$$N_{surp} = N_{in} - N_{out}, \tag{14}$$

The EUNEP framework defines a maximum amount of N_{surp} to have low environmental effect is below 80 kg/ha, but not below 0, as this implies N deficits that can deplete soil N reserves overtime, leading to reduced soil health and fertility (Giller, 2001). Additionally, a large N_{surp} was identified as the main reason for N leaching and N pollution across various locations (Klages et al., 2020; Chen et al., 2014; Zhou et al., 2016). We impose a maximum N_{surp} of 40 kg/ha, which we set in our defined reward function. This change is reflected in the NUE framework plot in Fig. 1. It plots N_{in} (x axis) against N_{out} (y axis). The white space in the figure is the target N efficiency, which emphasizes how difficult the task at hand is for the RL agent. We set this tighter constraint to discover RL policies that further reduce N_{surp}.

2.4.3. Reward function

In this section we define our reward functions and elaborate on our choices. From the perspective of crop management, the turnover of any management action is naturally delayed and sparse, i.e., an action's effect on the environment or yield can only be seen after the growing season. NUE and N_{surp} can only be calculated at the end of a growing season, inevitably running into the problem of very sparse signals. Reward function design requires many careful considerations: objective alignment, sparsity, simplicity, and shaping, among others (Sutton and Barto, 2018). It is an often underestimated aspect in many RL-application papers with specially designed reward functions. Moreover, Booth et al. (2023) found that many misdesigns in the reward function stems from mismatch perspectives of what the reward function communicates.

In essence, we take the advice of Sutton and Barto (2018), stating that "The reward signal is your way of communicating to the agent what you want achieved, not how you want it achieved". To that end, we design a reward function incorporating the EUNEP framework, specifically including NUE, N_{surp} and end-season yield in our objective. This reward signal is very sparse, so we scalarize this multi-variable reward function by designing a novel utility function (Rosenthal, 1985). The reward function we define as follows:

$$R = \phi N_{UE} \cdot \phi N_{surp} + Y_{\text{cond}}, \tag{15}$$

where R is given at harvest (*i.e.,* when an episode terminates). The terms ϕN_{UE} and ϕN_{surp} each return bounded signals with a range $[0, 1]$, depending on the values of NUE and N_{surp} of the episode. Y_{cond} is a term describing a conditional normalized yield with a range of $[0, 1]$. No penalty/punishment terms in the reward function were implemented. The reward function terms are described below:

$$\phi N_{\text{UE}} = \text{clip}\left(1 - \frac{|\text{NUE} - 0.7| - 0.2}{\omega_{\text{NUE}}}, 0, 1\right), \tag{16}$$

ϕN_{UE} is a clipped linear function. If NUE is in the range of [0.5, 0.9], the function returns 1. Here $\omega_{\text{NUE}} = 1$, which determines the width of the constraint. This wider range is a form of reward shaping to better guide the agent towards the desired NUE.

$$\phi N_{surp} = \text{clip}\left(1 - \frac{|N_{\text{surp}} - 20| - 20}{\omega_{\text{N}surp}}, 0, 1\right), \tag{17}$$

Similar to Eq. (16), ϕN_{surp} is a clipped linear function that returns 1 if N_{surp} is within a certain range. As explained beforehand, to give an agent a bigger incentive to reduce surplus we define the allowed range of N_{surp} to be [0, 40]. We set $\omega_{\text{N}surp}$ to be 100. Fig. 2 shows the shape

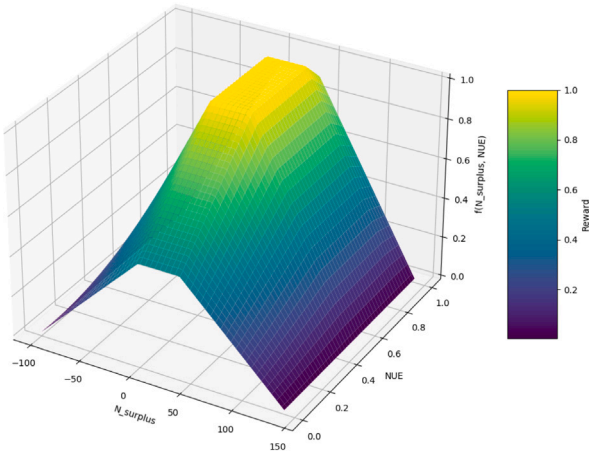

Fig. 2. A 3D plot that partially describes the reward function. x and y shows a range of N_{surp} and NUE, respectively, and z shows the output of the combinations.

of our designed reward function. In the yellow region (i.e., $z = 1$), the Y_{cond} term activates, providing additional reward (≥ 1) for higher yield. Further, Y_{cond} is defined as follows:

$$Y_{\text{cond}} = \begin{cases} \text{WSO}_{\text{norm}} & \text{if } \phi N_{\text{UE}} \cdot \phi N_{surp} = 1, \\ 0 & \text{otherwise, and} \end{cases} \tag{18}$$

$$\text{WSO}_{\text{norm}} = \frac{\text{WSO} - \text{WSO}_{\text{min}}}{\text{WSO}_{\text{max}} - \text{WSO}_{\text{min}}}, \tag{19}$$

We define Y_{cond} as a conditional normalized yield reward term. As shown in (18), this term added to the reward signal only if NUE and N_{surp} are within the desired ranges, otherwise this term does not affect the reward signal. The yield (*WSO*, Total Weight Storage Organ) is the simulated grain in WOFOST. The yield WSO_{norm} is normalized with min–max, where parameters for WSO_{min} we obtain from a low initial N setting with no fertilization events and WSO_{max} we obtain from the crop's average potential production throughout several years. Hence, it is possible to get a reward higher than 2; though the typical range of scalar reward that the agent can obtain is $[0, 2]$ with $R > 1$ indicating that the agent achieved the required NUE and N_{surp}.

Altogether, this reward signal was designed to encourage the agent to maximize yield only when it is within the environmental norms of the EUNEP framework. It is inevitable that NUE and N_{surp} produces a very sparse reward signal in our case. Thus, to guide the training process we implement an intrinsic reward signal (Pathak et al., 2017) to encourage exploration and introduce intermediate rewards for our agent. More details we elaborate on Section 2.5.3.

2.5. Experiments

We conduct experiments to answer our posed research questions. We evaluate how well our *NUE agent* performs against baselines, other agents trained with previously defined reward functions from literature to maximize yield within efficient NUE and N_{surp} ranges. Further, we evaluate our agent in several soil scenarios.

To answer RQ 1, we train the agents with WOFOST that was calibrated for the conditions in Lelystad, the Netherlands, and use winter wheat as the crop. We train our proposed model with the NUE reward function. This model we refer to as "*NUE agent*". We compare the performance of several constrained RL agents, trained with different objective/reward functions. These reward functions include *relative-yield*, *yield-N-loss*, and *financial*. Further, we evaluate the performance of *NUE agent* on our defined baselines. We describe our training setup further in Section 2.5.2.

To answer RQ 2, we evaluate *NUE agent* with several scenarios to test its adaptability to different soil conditions, comparing its performance against baselines. We compared two different scenarios:

1. different initial N content, and
2. different drought-sensitive soils.

A different initial N content at sowing date highly affects the outcome of NUE, as there needs to be adjustments regarding the amount of N fertilization to achieve a good balance of N. These initial nitrogen content reflect a range of pre-sowing conditions associated with different agricultural practices. For instance, low initial nitrogen levels may result from the harvest of crops with high nitrogen uptake (Tan et al., 2005), whereas high initial nitrogen levels could arise from enhanced mineralization following manure fertilization (Bouldin et al., 1984). These scenarios have been documented in previous studies (Huang et al., 2007).

The different soil profiles affect the N dynamics of the whole system, where we would like to compare performance with fast-draining soil and slow-draining soil. The fine and coarse soil texture scenarios selected in this study represents practical cases encountered in the Netherlands (Silva et al., 2021; Faber et al., 2021), making them well-suited for evaluating the potential of reinforcement learning in realistic agricultural management. We evaluated how well *NUE agent* adapts its fertilization policy against the changes in the soil dynamics to achieve good NUE.

In the following subsections, we describe in detail the testing conditions, training conditions, the baselines, intrinsic rewards, and the evaluations for reporting our results.

2.5.1. Testing conditions

In this section we describe the environmental conditions of where our agents and baselines are evaluated. Our testing location (i.e., soil and weather condition) is calibrated to Lelystad, the Netherlands. We use daily historical weather of years 1983 to 2021 ($n = 39$) while adhering to the N deposition amounts of the specific year (see Fig. C.6). This location has 7 soil layers, and the amount of initial nitrate ($NO_3 - N$) and ammonium ($NH_4^+ - N$) in the soil are 70 kg/ha and 0 kg/ha, respectively, indicating that all the ammonium has been converted to nitrate through nitrification prior to sowing. The soil texture is silty loam; medium coarseness. The parameters for the soil we show in Appendix A. We adopted the soil and site parameters from Berghuijs et al. (2024). In our experiments, we keep the CO_2 concentrations for every year fixed. We further discuss this choice in Section 4.

For the following experiments, we evaluate our agents in four different conditions. The high initial N scenario will have the same soil profile as the Lelystad conditions, but starts with 100 kg/ha of inorganic N content. The division of N for each soil layer was done as follows: 70% of the total inorganic N is deposited in the upper 30% of the soil and vice versa, consisting of 85% nitrate and 15% ammonium in total. The low initial N scenario starts with 5 kg/ha initial N content with a similar division of N as the high N scenario.

Next, the drought-sensitive soil (fast draining soil) has a sandy profile; a very coarse texture. On the other hand, the drought-insensitive soil (slow draining soil) has a clay profile; very fine texture. For the soil texture scenarios, we converted sample soil files from the original *fortran* WOFOST.[5] We defined certain parameters then convert it into a file readable by WOFOST. Appendix A further details this conversion process. The site parameters we kept the same as the first experiment.

2.5.2. Training conditions

We utilize randomization of certain aspects in the training pipeline to improve generalization of the RL agent (Tobin et al., 2017). We use random weather parameterized from climate variables obtained from the Lelystad weather station, hence we ensure the distribution of the generated weather closely matches the data. The soil profile follows the "PAGV" location defined in Groot and Verberne (1991). We also

[5] https://github.com/ajwdewit/WOFOST

employ randomization of initial N conditions, where we set a mean and standard deviation of 35 kg/ha and 15 kg/ha N, respectively. These values mirror the general variability of initial N conditions in the Netherlands. We also randomize the N deposition statistics for the calculation of NUE.

We train the *NUE agent* with 10 random seeds, for 3 million steps. The hyperparameters we use for training are listed in Table B.7. Further, we train several agents with different reward functions to see how well the obtained policies adhere to the NUE framework indicators. The agents have similar training conditions compared to *NUE agent* where only two things differ:

1. the reward function of each RL agent, and
2. which constraint functions are used in training.

We train *NUE agent* with our proposed NUE reward function, and constrain it with C_1 and C_2 (Eq. (2) and (3)). For the agents trained with the other reward functions, we directly constrain it with environmental indicators following Turchetta et al. (2022), constraining them with C_{1-4} (Eq. (2), (3), (4), (5)). In this case, C_3 and C_4 are functions that constrain NUE and N_{surp} within our defined efficient ranges.

The three agents we will compare *NUE agent* with are trained with the following reward functions:

1. *relative-yield* reward function, based on Overweg et al. (2021), Kallenberg et al. (2023);
2. *yield-N-loss* reward function, based on Wu et al. (2022), Tao et al. (2022); and
3. *financial* reward function, based on Turchetta et al. (2022).

relative-yield was trained with a reward function for optimizing yield by rewarding the agent based on the *additional* yield compared to a *zero fertilization* policy:

$$R_t = (WSO_t^{\pi} - WSO_{t-1}^{\pi}) - (WSO_t^{0} - WSO_{t-1}^{0}) - \beta N_t, \tag{20}$$

where N_t is the total amount of fertilizer applied, WSO^{π} is the grain growth with the agent's policy and WSO^0 is the growth in the zero fertilization policy. β is a multiplier for fertilization actions which we set to 10, following Kallenberg et al. (2023).

yield-N-loss was trained with a reward function that maximizes yield minimizing total N leaching:

$$R_t = \begin{cases} \omega_1 Y - \omega_2 N_t - \omega_3 N_{l,t} & \text{if episode terminate at } t, \\ -\omega_2 N_t - \omega_3 N_{l,t} & \text{otherwise,} \end{cases} \tag{21}$$

where ω_{1-3} are term weight modifiers that we set to 0.2, 1, and 5, respectively, following (Tao et al., 2022). $N_{l,t}$ is the amount of N leaching at time t. This reward function has a delayed positive reward, however the penalties are quite dense.

financial was trained with a reward function focusing on profitability for the farmer based on static prices of grain and N fertilizer:

$$R_t = €Y_t - €N_t, \tag{22}$$

where $€Y_t$ is a term that describes the price of winter wheat grain in $€/ha$ per time step, and $€N_t$ is the price of N fertilizer also in $€/ha$. The prices are based on the prices of winter wheat grain and N fertilizer in the Netherlands for the year 2020 (Wageningen Economic Research, 2023b,a): €181.67/1000 kg grain and €20.49/100 kg N fertilizer. It is possible to vary the prices for each year based on historical prices; however, this introduces unnecessary dependencies for the learning agent, as the learning agent does not know which year it is on. Hence, we keep the prices fixed.

2.5.3. Intrinsic rewards

In the case of inevitable sparse reward signals, a common method for improving the agent's learning process are intrinsic rewards, such as the Intrinsic Curiosity Module (Pathak et al., 2017). Intrinsic rewards are reward signals that are generated within the agent itself, rather than from the RL environment. It is a self-supervised method to encourage the agent to explore unfamiliar states, which helps the agent avoid converging to suboptimal behaviors.

The training conditions we employ for the agent can be categorized as *non-singleton* environments, i.e., environments where the training and testing conditions are different. In these type of settings, there is a risk for an agent to overfit on the training conditions (Zhang et al., 2018; Song et al., 2019). This hurdle is compounded with the addition of random initial N conditions, which potentially shifts the agent's target. To tackle that, we implement the Exploration via Elliptical Episodic Bonuses (E3B, Henaff et al., 2023), which is an intrinsic reward method that was developed to solve non-singleton RL environments that have randomized initial conditions/positions. Similar to the work of Pathak et al. (2017), it uses an inverse prediction model to predict whether a change in the environment was caused by the agent's action or not. On top of that, it introduces a count-method that gives scalar episodic bonuses to the agent when it sees a different initial condition. As it fits our problem setup, we implement E3B for the training of our RL agent to help it converge faster during training.

2.5.4. Baselines

To compare the performance of the RL agent, we implement two baseline agents:

1. the **Standard practice** agent (***N2***), and
2. the **Demeter** agent.

In this paper, the experimental conditions are calibrated following the conditions in the field experiments of Groot and Verberne (1991). Hence, this baseline represents the standard practice of farmers and a direct comparison to the results of previous literature. We define the standard practice agent following their "*N2*" fertilization regime in "PAGV", where they apply 3 different amounts each year in fixed dates. *N2* is a challenging-to-outperform baseline, as it is an optimal fixed amount of fertilization for the specific N initial condition in our test location and winter wheat variety.

Named after the Greek goddess of agriculture, the Demeter agent is an oracle/episode-optimized agent, where timing and amounts of fertilization are optimized for the year it is evaluated on. Different from the *Ceres* agent from Kallenberg et al. (2023), it divides its application of N to different times in the growing season, ensuring time-wise optimality for fertilization actions. Unlike an RL agent, the Demeter agent can observe future weather and hence serves as an upper-bound for fertilization actions. Similar to the RL agent, we constrain its cumulative fertilization actions to 3 for each season to keep a fair comparison. We use a function optimizer: Generalized Simulated Annealing (Bohachevsky et al., 1986), to determine the correct amount of fertilizer needed for each week.

2.5.5. Evaluation

We mainly evaluate our RL agents with yield, NUE, N_{surp} and cumulative reward indicators. Also we show cumulative fertilization amounts, N loss and profit. By showing profit obtained, we can see whether optimizing NUE affects profit. The profit reports are based on prices from 2020. N loss reported is a combination of N leaching and denitrification loss. We evaluate through reporting the obtained medians and 95% confidence interval throughout all test years and random seeds, without explicitly handling any outliers to ensure good representation of results. Further, the one-sided p-values are aligned appropriately to the direction of each metric. Next, we plot the best model and baselines on the NUE framework graph. Additionally, we evaluate how many years the agents did well in reaching the target NUE and N_{surp}.

Table 2
Results for target indicators. We report the medians (and 95% bootstrapped confidence intervals). Some icons describe the target values of the indicators: ↑ indicates higher is better, ↓ indicates lower is better and ↔ indicates a target between two values.

Agent	Yield [tons/ha] (↑)	N_{surp} [kg/ha] (0↔40)	NUE [–] (0.5↔0.9)	Reward [–] (↑)
NUE Agent	9.43 (8.97, 9.65)	27.2 (21.0, 34.0)	0.87 (0.84, 0.90)	0.99 (0.96, 1.93)
relative-yield	9.61 (9.26, 9.81)	50.2 (41.0, 54.0)	0.79 (0.77, 0.81)	0.85 (0.81, 0.93)
yield-N-loss	9.05 (8.53, 9.34)	−11.1 (−27.3, 11.9)	1.07 (0.93, 1.16)	0.56 (0.35, 0.71)
financial	9.93 (9.54, 10.08)	120.4 (88.2, 181.2)	0.61 (0.53, 0.67)	0.19 (0.0, 0.51)
Demeter	9.73 (9.49, 9.89)	39.9 (39.9, 39.9)	0.83 (0.82, 0.83)	2.05 (1.99, 2.09)
N2	9.61 (9.54, 9.88)	52.6 (49.9, 59.4)	0.78 (0.76, 0.80)	0.86 (0.80, 0.90)
$\Delta_{\text{NUEAgent,N2}}$	−0.15 (−0.20, −0.09) $p = 0.0001$	−27.4 (−28.9, −25.7) $p = 1e-6$	+0.09 (0.08, 0.10) $p = 0.0$	+0.39 (0.08, 1.09) $p = 0.0011$

Table 3
Results for additional indicators.

Agent	Fertilization [kg/ha] (↓)	N Loss [kg/ha] (↓)	Profit [k€/ha] (↑)
NUE Agent	190.0 (180.0, 190.0)	42.6 (38.2, 49.2)	1.44 (1.37, 1.48)
relative-yield	210.0 (210.0, 220.0)	47.9 (41.0, 47.9)	1.46 (1.42, 1.50)
yield-N-loss	140.0 (130.0, 160.0)	34.8 (30.7, 42.6)	1.39 (1.31, 1.44)
financial	290.0 (240.0, 350.0)	46.3 (39.6, 55.8)	1.49 (1.42, 1.52)
Demeter	202.3 (193.0, 210.6)	37.6 (35.1, 43.5)	1.49 (1.45, 1.51)
N2	220.0 (220.0, 220.0)	45.0 (40.5, 50.2)	1.49 (1.46, 1.49)
$\Delta_{\text{NUEAgent,N2}}$	−30.0 (−30.0, −30.0) $p = 1e-5$	−3.9 (−3.2, 5.4) $p = 0.3028$	−0.01 (−0.06, 0.03) $p = 0.9871$

3. Results and analysis

In this section we report quantitative results of our experiments, mainly answering the imposed research questions (Sections 3.1 and 3.2). Further, we qualitatively compare the results to empirical findings of previous work related to NUE (Section 3.3).

3.1. Performance of agents

In this section, we compare our *NUE agent* with baseline methods and other RL agents (trained with *LagPPO*) in the Lelystad case study to address research question 1. Results are summarized in two tables: Table 2 reports the median yield, N_{surp}, NUE, and reward for each agent, and Table 3 reports the cumulative fertilization, N loss, and profit. Additionally, we report the RL training curves in Fig. C.9 in the appendix.

Higher N_{surp} is needed to maximize yield, as shown by *Demeter*, which keeps N_{surp} near 40 kg/ha. Among the agents and *N2*, *NUE agent* achieved the highest median reward, and the *relative-yield* agent performed similarly. These results indicate that a reward based on additional yield relative to a zero-N treatment can improve NUE and N_{surp}.

Agent *yield-N-loss* obtained policies that reduced N Loss by applying considerably less fertilizer. However, this is at the cost of negative N_{surp}, and high NUE, indicating N imbalances and soil N removal/mining.

Agent *financial* increases profit at the expense of N_{surp} and higher N loss, leading to the lowest reward. Although we applied a Lagrangian constraint for NUE and N_{surp}, the constraint critic did not predict these well (see Fig. C.7). This signifies the difficulty of directly constraining NUE and N_{surp}.

Demeter shows that good profit is possible with good NUE. The *financial* agent shows a large gap in N_{surp}, as reflected in its fertilization, but the difference in N loss is small. This is because N_{surp} is defined as N input minus N in the grains, while the crop allocates some of this N to vegetative organs. This exhibits that *NUE agent* can discover the optimal amount of fertilization that improves NUE. The subsequent scenario experiments (Section 3.2) further reveal the significant impact of different soil profiles on N loss.

The differences between *NUE agent* and *N2* are detailed in Tables 2 and 3. *NUE agent* achieves similar yield with lower N_{surp}, N loss, and fertilizer use, which leads to lower profit than *N2*. Only *NUE agent* meets the N_{surp} target, which lowers environmental risk and preserves N balance (Klages et al., 2020). This shows that *NUE agent* can discover policies that ensure efficient NUE with low N_{surp}.

Fig. 3 shows the NUE framework for each agent. *Demeter* meets the target by staying near the N_{surp} or NUE boundary. *NUE agent*, *N2*, and *relative-yield* perform similarly, though *NUE agent* and *relative-yield* have more years within the efficient range. In contrast, *yield-N-loss* and *financial* agents have a wider spread. as confirmed by the kernel density estimate. For instance, the *financial* agent has only 19 years with N inputs below 300 kgN/ha/y, while the *yield-N-loss* agent sometimes enters the soil mining region.

Fig. 4 presents the fertilization actions of *NUE agent*, the *relative-yield* agent, *N2*, and *Demeter* in relation to precipitation during 2020.

A limitation of the RL approach is that *NUE agent*, *N2*, and *relative-yield* show similar scatter patterns with small shifts along the N input axis. Notably, *NUE agent* applies less fertilizer than *N2* and *relative-yield* (Table 3). Each NUE plot in Fig. 3 contains a single outlier year — with an N output of approximately 140 kg/ha — corresponding to the year 2020. An extreme precipitation event occurred on the flowering date (Fig. 4), which inhibited N uptake (Kowalenko and Bittman, 2000). Because the RL agent makes decisions based solely on immediate observations, it cumulatively applied 190 kg/ha in preceding timesteps, failing to anticipate the disturbance. Moreover, Fig. 4 illustrates that the *LagrangianPPO* algorithm satisfied constraints C_1 and C_2 (Eqs. (2) and (3)) for both the *NUE agent* and the *relative-yield* agent.

Fig. 5 presents box plots that more intuitively display the spread of NUE and N_{surp} for each agent; the legend indicates the number of test years meeting the target requirements. For NUE, *N2* and *relative-yield* meet the target in all test years, while *NUE agent* has 10 years in the soil mining region. In terms of N_{surp}, *NUE agent* performs best. The *relative-yield* agent has a median N_{surp} of 50 kgN/ha, and the others do not consistently meet the target. Fig. C.8 shows individual years for each agent.

3.2. Performance in different soil scenarios

In this section, we present the results of the soil scenario experiments detailed in Tables 4 and 5. We report each agent's performance under different scenarios, including *NUE agent* and the baselines, and we also evaluate the *relative-yield* agent given its competitive performance in the previous experiment. *Demeter* provides the optimal

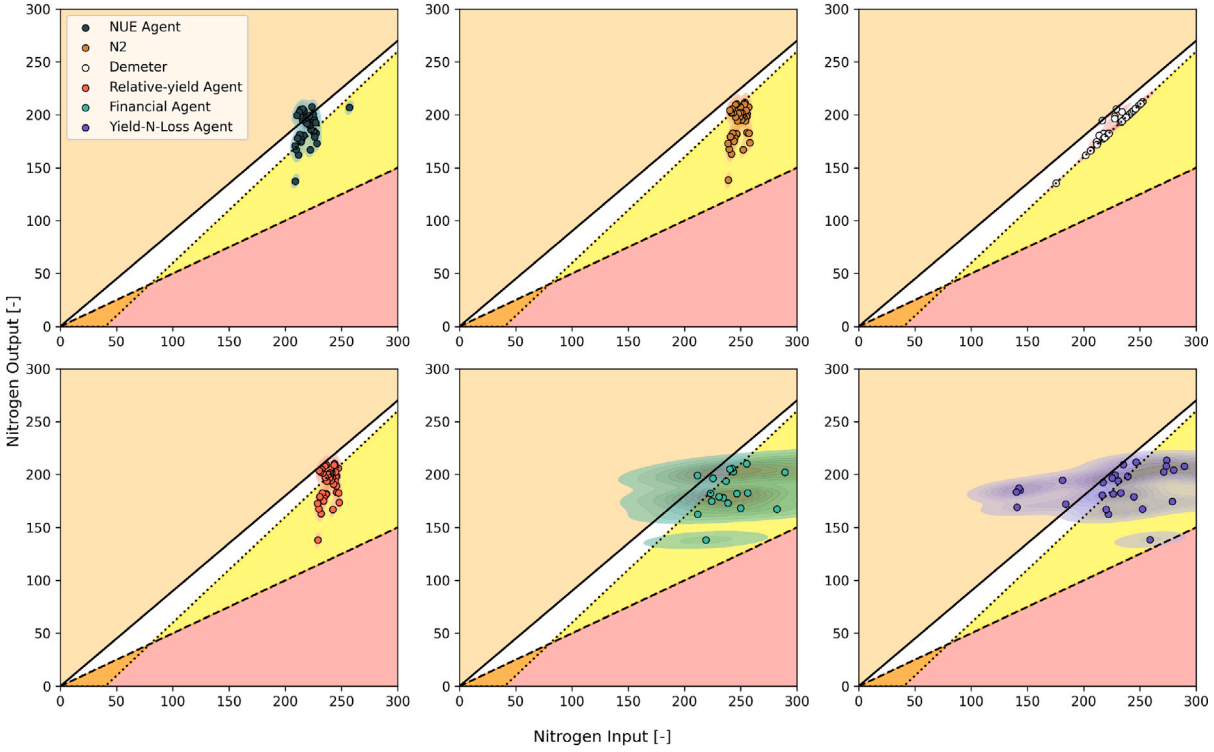

Fig. 3. NUE graphs for each agent. Each dot represents a single test year ($n = 39$). A kernel density estimate (KDE) was applied around plotted to show the spread of performance.

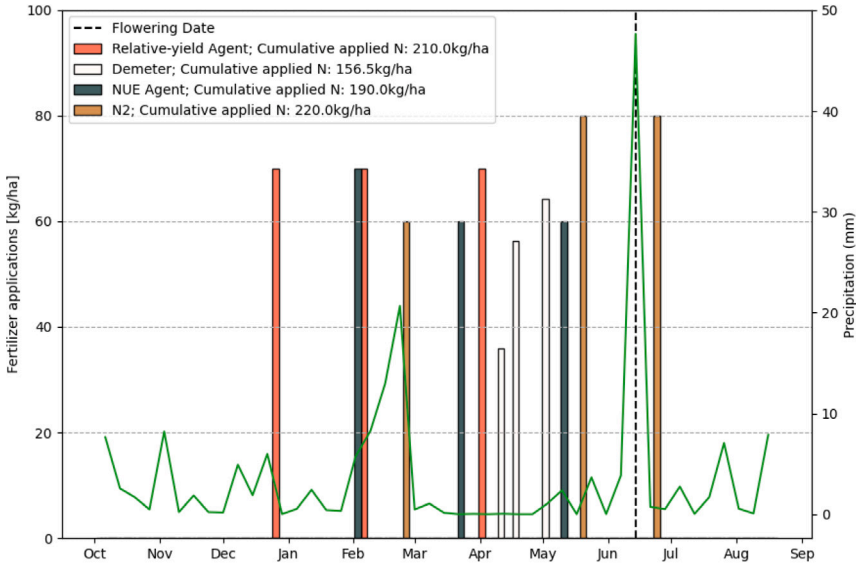

Fig. 4. A plot showing fertilization actions of *NUE agent*, *relative-yield* agent, *N2* and *Demeter*. The dotted line shows the flowering date.

metrics for each scenario. Our primary aim here is to address research question 2.

For low and high initial N scenarios, *NUE agent* maintains performance similar to normal conditions with similar median rewards. *N2* also performs consistently despite the change in initial soil N. *Demeter* applied more fertilization in the high-N scenario than in the low-N scenario, likely because lower initial N limits early growth and reduces subsequent N demand. This observation further suggests that the initial N content does not alter the target N scenario, as higher initial N is largely lost through leaching early in the growing season. *NUE agent* reliably applies the target cumulative N, and the *relative-yield* agent performs similarly.

In soil type scenarios, *NUE agent* generally meets the target NUE and N_{surp} in fine soil, though with some variability. Nevertheless, *NUE agent* delivered higher yield than the *relative-yield* agent while using less fertilizer. All agents show low N loss in fine soils due to reduced leaching.

In the coarse soil scenario, prone to drought and leaching, *NUE agent* outperforms *N2* by achieving lower N loss (95.7 kg/ha vs. 101.3 kg/ha) and better N_{surp} (54.9 kg/ha vs. 78.3 kg/ha), although N_{surp} remains above target. Both *NUE agent* and *relative-yield* suffer yield reductions (7.73 and 7.91 tons/ha, respectively), highlighting the challenge of maintaining productivity in such soils.

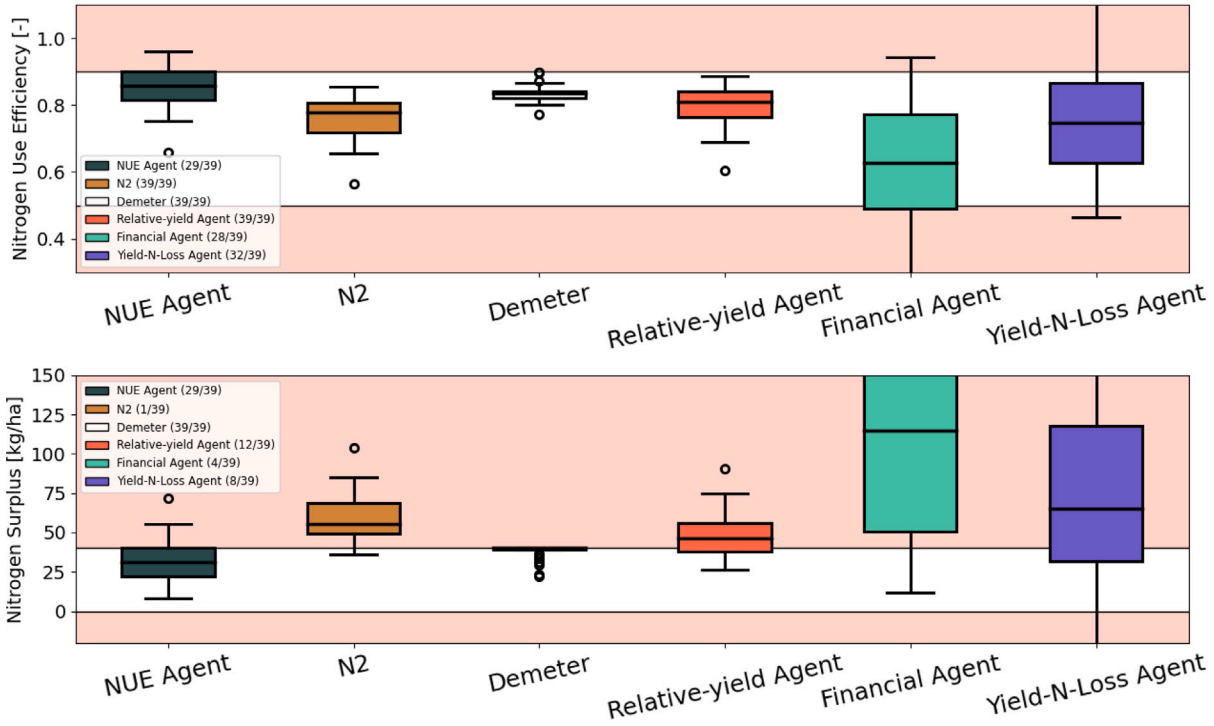

Fig. 5. Figure showing box plots of years within efficient N. The legend details how many years out of the total test years that are within the NUE (top) and N_{surp} (bottom) requirements.

Table 4
Results showing target indicators for each scenario.

Agent	Scenario	Yield [tons/ha] (↑)	N_{surp} [kg/ha] (0↔40)	NUE [–] (0.5↔0.9)	Reward [–] (↑)
NUE Agent	*Low N*	9.19 (8.78, 9.41)	32.0 (25.2, 39.7)	0.85 (0.81, 0.88)	0.98 (0.93, 1.69)
	High N	9.40 (9.29, 9.46)	28.9 (21.2, 37.1)	0.87 (0.83, 0.90)	0.97 (0.92, 1.68)
	Fine Soil	9.88 (9.38, 10.03)	24.7 (12.4, 67.0)	0.88 (0.74, 0.93)	0.88 (0.28, 0.90)
	Coarse Soil	7.73 (7.04, 8.63)	54.9 (35.3, 76.0)	0.74 (0.67, 0.81)	0.76 (0.50, 0.82)
relative-yield	*Low N*	9.55 (9.21, 9.73)	64.9 (57.4, 73.1)	0.74 (0.72, 0.77)	0.75 (0.66, 0.82)
	High N	9.54 (9.01, 9.74)	43.8 (37.5, 52.8)	0.81 (0.79, 0.84)	0.89 (0.83, 0.96)
	Fine Soil	9.44 (8.53, 10.02)	36.7 (20.3, 49.6)	0.85 (0.80, 0.91)	0.76 (0.56, 0.89)
	Coarse Soil	7.91 (7.31, 8.57)	70.0 (54.5, 83.5)	0.69 (0.64, 0.76)	0.67 (0.50, 0.83)
Demeter	*Low N*	9.47 (9.29, 9.67)	39.9 (39.9, 39.9)	0.83 (0.82, 0.83)	1.99 (1.94, 2.04)
	High N	9.76 (9.52, 9.88)	39.9 (39.9, 39.9)	0.83 (0.82, 0.83)	2.05 (2.00, 2.09)
	Fine Soil	10.01 (9.97, 10.06)	39.9 (39.9, 39.9)	0.84 (0.83, 0.84)	2.12 (2.06, 2.13)
	Coarse Soil	8.57 (7.55, 9.19)	39.9 (39.9, 39.9)	0.81 (0.80, 0.82)	1.73 (1.92, 1.51)
N2	*Low N*	9.59 (9.40, 9.72)	57.4 (52.6, 61.8)	0.77 (0.75, 0.79)	0.82 (0.78, 0.87)
	High N	9.72 (9.06, 9.87)	54.7 (49.4, 60.2)	0.78 (0.75, 0.80)	0.85 (0.79, 0.90)
	Fine Soil	10.04 (9.80, 10.07)	46.3 (42.5, 51.5)	0.81 (0.79, 0.82)	0.93 (0.88, 0.97)
	Coarse Soil	8.54 (7.63, 8.93)	78.3 (68.7, 91.3)	0.69 (0.63, 0.72)	0.62 (0.48, 0.71)

3.3. Results compared to previous literature

None of the previous works in RL for crop management evaluated the NUE performance of their experiments, prohibiting direct comparisons. In this section, we investigate non-RL literature that describes results relevant to this work and quantitatively analyze their findings to enhance the context of our findings.

In preliminary work, Silva et al. (2018) analyzed the NUE performance of farms in the Netherlands for winter wheat. They found more than half of the fields had N_{surp} of more than 80 kg/ha and 40% were outside the desirable range of NUE. In follow-up work, Silva et al. (2021) assessed NUE performance for a large database of farmers in the Netherlands with the EUNEP framework. They found that the NUE for winter wheat was roughly 0.8 and N_{surp} was roughly 78 kg/ha, which matches the results of *N2* in our experiments. Moreover, when comparing between coarse and fine soils for winter wheat, they found that coarse soils increase N_{surp} compared to fine soils, while the opposite is true for NUE. These results corroborate our findings in the soil scenario experiments (shown in Table 4 for the *N2* agent) and further adds evidence that RL is capable of improving these metrics through better fertilization policies.

Next, Faber et al. (2021) investigated N fertilization regimes to reduce N losses in light/coarse soils for 2300 farms cultivating winter wheat in Poland. They found that the coarse soils generally reduced NUE and made it difficult to achieve profit. Interestingly, they indicate that a strategy that keeps the farms profitable generally require farms to soil mine. In our experiments, the coarse soils indeed reduced profit due to inefficient N uptake, which still has potential to be optimized as shown by the *Demeter* agent (Table 5).

A research conducted by Ravensbergen et al. (2024) on NUE performance of ware potato under coarse and fine soils in the Netherlands reveal high N_{surp} variability without high variability on yield. They further found that some cases with higher yields had relatively low N_{surp}. This study concludes that it is possible to reduce N inputs while still maintaining higher yields through optimized timing and

Table 5
Results showing additional indicators for each scenario.

Agent	Scenario	Fertilization [kg/ha] (↓)	N Loss [kg/ha] (↓)	Profit [k€/ha] (↑)
NUE Agent	*Low N*	190.0 (190.0, 210.0)	11.7 (7.5, 11.7)	1.41 (1.34, 1.44)
	High N	190.0 (180.0, 190.0)	72.8 (63.8, 81.3)	1.44 (1.37, 1.48)
	Fine Soil	190.0 (190.0, 240.0)	0.4 (0.2, 3.4)	1.49 (1.41, 1.53)
	Coarse Soil	190.0 (160.0, 200.0)	95.7 (80.5, 95.7)	1.19 (1.07, 1.27)
relative-yield	*Low N*	230.0 (220.0, 240.0)	12.3 (9.5, 16.3)	1.45 (1.42, 1.47)
	High N	210.0 (200.0, 210.0)	81.4 (73.9, 88.4)	1.46 (1.38, 1.49)
	Fine Soil	210.0 (210.0, 210.0)	0.3 (0.3, 1.8)	1.47 (1.31, 1.53)
	Coarse Soil	210.0 (210.0, 210.0)	109.8 (91.8, 123.1)	1.21 (1.19, 1.30)
Demeter	*Low N*	199.6 (190.0, 204.8)	5.6 (4.6, 6.1)	1.45 (1.42, 1.48)
	High N	205.2 (199.9, 210.9)	70.0 (67.0, 77.6)	1.49 (1.46, 1.52)
	Fine Soil	212.4 (208.0, 217.0)	0.4 (0.3, 1.9)	1.53 (1.49, 1.54)
	Coarse Soil	180.9 (166.5, 187.3)	59.3 (55.9, 63.1)	1.31 (1.15, 1.41)
N2	*Low N*	220.0 (220.0, 220.0)	12.6 (10.8, 15.1)	1.47 (1.43, 1.48)
	High N	220.0 (220.0, 220.0)	79.2 (72.6, 90.2)	1.49 (1.37, 1.51)
	Fine Soil	220.0 (220.0, 220.0)	0.8 (0.2, 1.7)	1.53 (1.49, 1.54)
	Coarse Soil	220.0 (220.0, 220.0)	101.3 (93.5, 110.8)	1.30 (1.15, 1.36)

amount of N inputs, which is a task that we seek to solve through recommendations of RL.

4. Discussion, limitations and future work

In this section, we delve deeper into the design choices, assumptions, and experimental findings that shaped our approach. We then discuss the limitations of our work and propose directions for future research.

4.1. Main discussions

Reward function. We formulated a novel NUE reward function because we aimed to optimize three indicators simultaneously (yield, NUE and N_{surp}). This formulation may benefit the use of the multi-objective RL framework (Hayes et al., 2022). However, we do not use this framework and instead scalarize the multiple objectives through our designed reward function, using a utility function approach (Wierzbicki, 1980). While our results show that this works well in guiding the RL agent to our multi-objective target, it is possible that a better objective or reward function formulation exists.

Observation space. In this work, we have an observation space that is fairly complete, including some hard-to-measure features such as soil and deposition observations for NO_3 and NH_4, total N grain content, crop transpiration rate, N loss and crop N uptake. In reality, these crop features are difficult to measure and will incur costs and labor to measure. Hence, it is infeasible for an RL agent to observe all of these features in every timestep.

Constraints. We impose constraints on fertilization frequency and crop development stage to ensure that RL recommendations are actionable and to prevent the agent from "reward hacking" the NUE target. Early experiments revealed that agents would wait until the end of the season to reduce N input (increasing NUE) and then fertilize before harvest—a practice that is unrealistic and potentially harmful. While our constraints addresses this issue and follows safety learning protocols (Ji et al., 2023), it introduces a Pareto trade-off in the achievable rewards (Censor, 1977). Although daily small-dose fertilization would optimize N uptake (Guertal, 2009), it is impractical due to fixed costs such as labor and machinery. Consequently, in this study, we do not use the standard PPO, since we require all our crop management decisions to be constrained to be actionable for practical use cases. In Fig. 4, we show that the actions of the RL agents are constrained and actionable, *i.e.*, not frequent and spaced out throughout the growing season. The work Turchetta et al. (2022), Kallenberg et al. (2023) used standard PPO, which entails frequent fertilization actions that are not feasible for a farmer.

Next, Turchetta et al. (2022) suggest directly constraining environmental effects such as N leaching or emissions for RL agents. We agree that this is essential. In our work, we constrain NUE and N_{surp} via Eqs. (4) and (5) rather than directly constraining N loss. However, our experiments reveal that the constraint critic fails to accurately predict NUE and N_{surp}, even with full access to N input and output data. This indicates that constraining these composite metrics is challenging. Therefore, incorporating NUE and N_{surp} directly into the reward function (as with *NUE agent*) provides a stronger learning signal for the agent.

Future improvements may involve alternative function approximators, pre-training the networks, or adopting hard-constraint algorithms such as Constrained Policy Optimization (CPO, Achiam et al., 2017). Further, to consider a holistic and unified approach, future work will include the exploration of different RL methods (e.g., value-function methods, model-based methods, different architectures such as RNNs or Transformers) to address various shortcomings of our current approach (for instance imitation learning, or offline learning). Consequently, including constraints in these RL methods are not trivial, and requires in-depth research in the domain of safe-learning (Ji et al., 2023).

Broader study case. NUE and N_{surp} is a common problem in the Netherlands. Hence, our experiments were done with a representative Dutch case study. Extending this problem to other cases requires a simple retraining of the RL agent with a WOFOST calibrated to a different location. Moreover, Kallenberg et al. (2023) has explored this idea and shown that the RL agent is robust when deployed in a location with a different climate. Notwithstanding, differing soil profiles remain a challenge.

Relevant literature in RL for crop management. In this section we discuss additional studies in reinforcement learning for crop management that were not previously mentioned. The work of Maillard et al. (2023) introduces *Farm-gym*, a customizable RL environment for crop management that models farms management as a dynamic system with multiple interacting elements. This environment supports tasks such as fertilization, irrigation, and pesticide application, and incorporates different soil scenarios (clay and sandy soils), which is directly relevant for enhancing NUE and N_{surp}. Next, Chen et al. (2023) investigated the use of RL in cotton irrigation. In their experiments, they incorporated soil characteristics in the DSSAT model, and then compared the performance of the RL agent to real world trials, which resulted in a more accurate simulation. The RL agent performed better in discovering optimal irrigation policies. The authors achieve similar results in a follow up work (Chen et al., 2025).

Soil profiles. Our experiments suggest that the RL agents found it difficult to generalize well to these different soil profiles. These soil profiles influence the changes in water and N, which exhibit strong temporal dynamics throughout the growing season and are unique to each soil profile. The influence of soil types in crop management is well supported by the literature (Silva et al., 2021; Faber et al., 2021; Wang et al., 2020; Raza and Farmaha, 2022; Ye et al., 2024), which further highlights the importance of considering soil dynamics in research pertaining to data-driven crop management recommendations.

Simulation-to-reality gap. The experiments conducted in this paper were done in-silico with a calibrated crop model. Although this simulated environment allows for rigorous testing, a significant simulation-to-reality gap remains. It is still an open question how well the RL agent would fare in the real-world, since there is a simulation gap between WOFOST and the actual farm. We argue that we should develop the RL approach in-silico before bringing it to the real-world, for the ability to rigorously test different methods. Moreover, RL methods that are mature and robust will consequently lower the barrier for technological adoption for the field practitioners. Nonetheless, moving towards bringing an RL-based recommendation tool to field trials should be a priority and the ultimate step to bring this research to maturation.

4.2. Limitations and future work

Our study has several limitations which are implicitly discussed in the previous section. In this section, we explicitly detail each of these limitations and suggest follow up work to potentially address these shortcomings.

1. ***NUE Reward Function:*** A limitation to our reward approach is the scalarized reward function, which may not fully capture the complex trade-offs between yield, NUE, and N_{surp}. Alternative multi-objective formulations might lead to the discovery of more balanced policies. Future work could explore the use of the Multi-Objective RL (MORL, Roijers et al., 2013) and, consequently, algorithms that excel in optimizing MORL problems. Another opportunity for a better reward function formulation is by using inverse RL (IRL, Arora and Doshi, 2021) to discover the objective of an optimal agent, which could be an expert farmer or the *Demeter* agent.
2. ***Observation Space:*** The extensive observation space, while beneficial in simulation, includes features that are impractical to measure in the field, potentially limiting real-world deployment. We suggest future work to employ imitation learning (Tao et al., 2022) or include measuring as part of the decisions (Baja et al., 2025), in the context of optimizing NUE.
3. ***Generalization to Diverse Soils:*** A limitation of our RL approach is that training on a single representative soil profile led to poor performance on different soil textures. This highlights the challenge and effect of soil profiles on effective fertilization policies. From an RL perspective, training with a wide range of soil profiles or training with perturbations/noise in the soil parameters could ensure a good exposition to a wide range of soil dynamic responses, and facilitate learning to generalize in diverse soil conditions (Tobin et al., 2017). Nonetheless, perturbations in soil parameters could also mitigate the RL bias towards the simulator when deploying in the real-world. We propose future work to focus on mitigating this issue through the use of RNNs (e.g., LSTMs or GRUs) or appending observations from preceding timesteps to capture the temporal dynamics of each soil type.
4. ***Handling Extreme Events:*** Our approach failed in extreme weather events, as detailed in Section 3.1. To address this, we propose a few approaches for future work:
   (a) employing RNNs to capture temporal dynamics of the soil, which might implicitly capture a change in the weather dynamics;
   (b) adopting a model-based RL approach to forecast future conditions;
   (c) integrating weather forecasts into the observation space.

   Nonetheless, predicting extreme events remain a formidable challenge (Camps-Valls et al., 2025), and requires specialized algorithms to predict.
5. ***Simulation-to-Reality Gap:*** Arguably, this point is the biggest limitation in our experiments. Our approach has been evaluated in-silico using WOFOST. However, since we use simulated data for the experiments, the performance of the RL agents in real-world conditions remains uncertain. Future research should focus on bringing the recommendations of the RL agent to in-field experiments. There are several established methods to deal with transferring RL simulations to reality:
   (a) deploy the simulator (WOFOST) and a trained RL agent in the real-world through a digital twin (Pylianidis et al., 2021) and data assimilation (Gaso et al., 2023).
   (b) train the RL agent with *offline* data from a real farm (Zhou et al., 2023), which reflects the true dynamics of the farm, ensuring accurate policy learning with virtually no simulation-to-reality gap;
   (c) employ RL agents that are robust towards distribution shifts (Luo et al., 2024), which ensures the agent is not biased towards the simulator's dynamics when employed in the real-world;
   (d) transfer learning and fine-tuning to a real-farm data (Taylor and Stone, 2009), which entails pre-training an agent in a simulator (e.g., WOFOST), and then fine-tuning with offline data from the farm.

   In the future, we will work on the digital twin approach.

5. Conclusion

In this work, we explored the potential of RL to optimize NUE in simulated crop management by introducing a novel reward function that balances yield, N_{surp}, and NUE with practical action constraints. We conducted two experiments: one comparing an RL agent trained with a NUE-oriented reward to baseline practices and alternative agents, and another assessing its robustness across varying soil conditions. Our results show that the NUE RL agent achieves optimal NUE and N_{surp} levels — reducing nitrogen surplus compared to standard practices — while remaining robust to shifts in initial soil N, though it faces challenges with extreme soil textures. These findings underscore the importance of considering both environmental and practical constraints when translating RL-based fertilization policies to real fields.

RL offers a compelling framework for developing adaptive fertilization policies that respect both agronomic and environmental objectives. By incorporating realistic agronomic constraints (e.g., limited fertilization events), the learned policies become more actionable and more likely to be adopted by farmers. Nonetheless, simulated data limits real-world applicability. Based on our findings, realizing a deployable RL-based fertilizer recommendation system to the real-world requires a holistic approach, incorporating different aspects of RL. In this paper, we discussed — from a unified RL and crop management perspective — the challenges faced, limitations of the paper and suggested follow-up work to pursue. We argue, before bringing these experiments to the real-world, the RL methods must comply with the expectation of the field practitioners in order to minimize the gap for technological adoption.

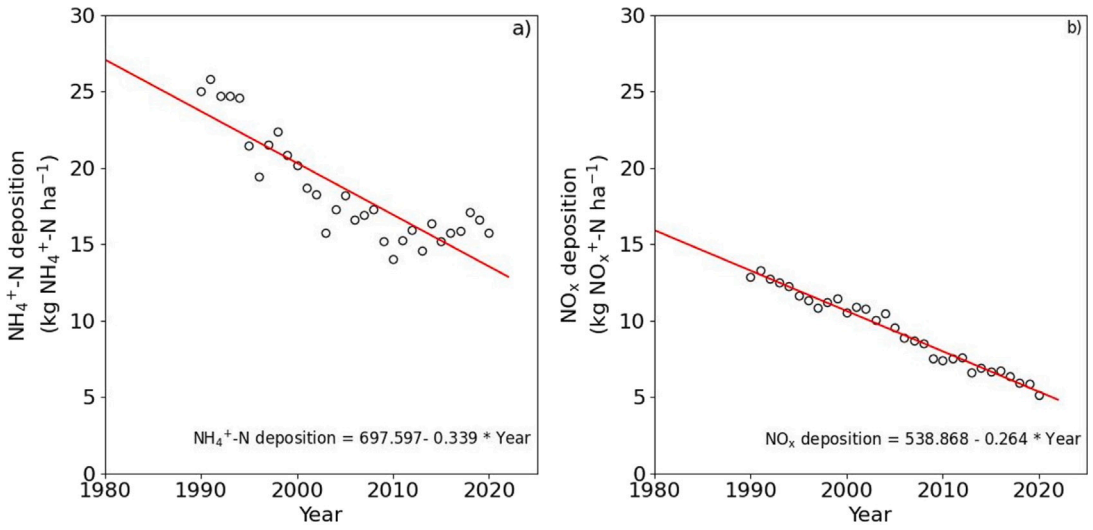

Fig. C.6. Trends of annual depositions of NH_4^+-N and NO_x-N in the Netherlands, taken from Berghuijs et al. (2024).

Table A.6
Soil parameters used for the experiments. We define soil that has 7 layers.

Parameter	Value
PFFieldCapacity	2.0
PFWiltingPoint	4.2
SurfaceConductivity	75.0
Thickness	[20.0, 10.0, 10.0, 20.0, 20.0, 20.0, 25.0]
CNRatioSOMI	[9.0, 11.0, 12.6, 14.31, 16.42, 18.0, 18.0]
FSOMI	[0.02, 0.015, 0.011, 0.0076, 0.0038, 0.001, 0.001]
RHOD	[1.406, 1.420, 1.432, 1.45, 1.505, 1.537, 1.537]
Soil_pH	[7.4, 7.4, 7.4, 7.4, 7.4, 7.4, 7.4]

Table B.7
LagrangianPPO hyperparameters.

Hyperparameter	Values
Learning Rate	1e−3
Batch Size	276
Gamma (γ)	1
Clip Range	0.2
GAE Lambda (λ)	0.95
Epochs	10
Value Function Coefficient (c_v)	0.5
Entropy Coefficient (c_e)	0.01
Max Gradient Norm	0.5
Timesteps per Update	2208
Policy Architecture	MLP
Activation Function	Tanh

A core motivation for this work is to bridge the existing gap between two communities: ML or RL researchers tend to focus on algorithmic innovation without sufficient agronomic input, while agronomists may find purely ML- or RL-driven papers too theoretical or detached from on-farm realities. By aligning performance metrics with agro-environmental indicators (NUE, N_{surp}) and structuring constraints around realistic field practices, our study bridges the gap for more effective collaboration between both communities. To open further development and collaboration, we provide documentation and code[6] of *CropGym*. By jointly engaging with agronomy experts, policymakers, and farmers, RL has the potential to evolve from an intriguing computational tool into a practical engine for global food security and environmental stewardship.

[6] https://github.com/WUR-AI/NUE_PCSE-Gym

CRediT authorship contribution statement

Hilmy Baja: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Michiel G.J. Kallenberg:** Writing – review & editing, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Herman N.C. Berghuijs:** Writing – review & editing, Software, Resources, Methodology. **Ioannis N. Athanasiadis:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the European Union Horizon Research and Innovation programme, Smart Droplets (Project Code: 101070496).

Appendix A. Conversion of soil parameters for soil texture experiments

We converted original CABO sample soil files of fortran WOFOST to a *.yaml* format that is readable by WOFOST SNOMIN. We chose specifically the files *ec1.CABO* and *ec6.CABO* that are soil files that contain parameters for coarse and fine soil types, respectively. We first define a specific set of parameters required by WOFOST SNOMIN, such as surface conductivity and thickness of each soil layer, which we show in Table A.6. Next, for each soil layer, we plug-in parameters from the *.CABO* files, namely: *CRAIRC, SMTAB, CONTAB*, which are the critical soil content for aeration, soil moisture content table, and 10-log hydraulic conductivity table, respectively. In general, only three parameters we change with respect to the soil file in the first experiment.

Appendix B. LagrangianPPO

All the RL agents trained in this study uses the LagrangianPPO algorithm, and we report the hyperparameters used during training in Table B.7.

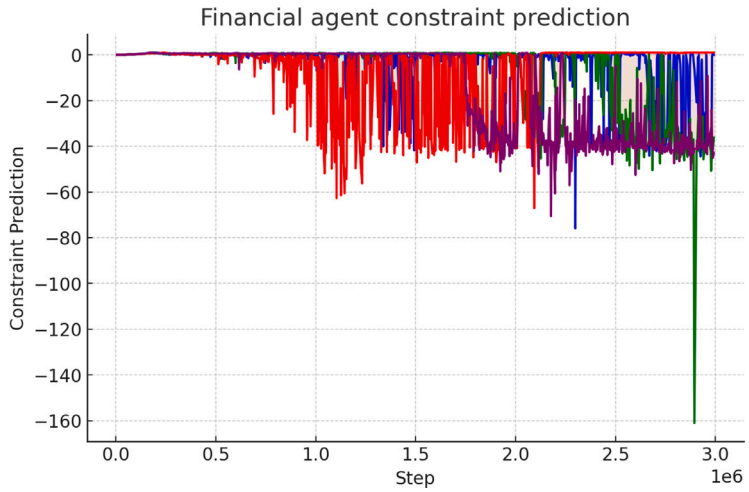

Fig. C.7. Prediction of the constraint network for the *financial* agent for several seeds. Here, the explained variance of the constraint predictions are shown. It is calculated as $var(y - y_{pred})/var(y)$, where the variance is $var = (\Sigma_i |a_i - \bar{a}|^2)/N$, where a is the array of constraint values and N is the length of the array. Here, we show that the constraint networks struggle to predict the constraint functions of C_3 and C_4, highlighting the difficulty of directly constraining NUE and N_{surp}.

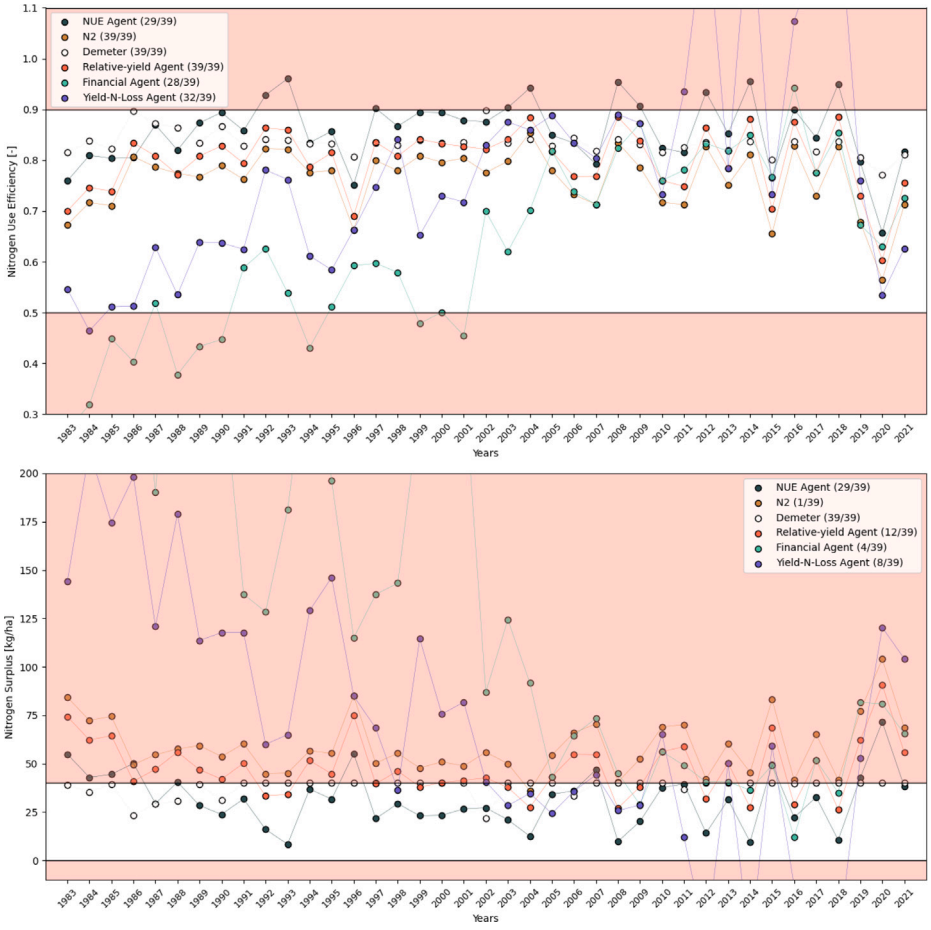

Fig. C.8. Scatter plot similar to Fig. 5, individual years for all the agents. The lines between years do not depict any trend and are only there for visual support.

Appendix C. Figures and tables

In this section we show several figures that are referenced in the main text. Specifically, we include (i) N deposition trend in Fig. C.6, (ii) constraint network predictions of RL agents in Fig. C.7, (iii) the NUE and N_{surp} performance of each agent in each year in Fig. C.8, and (iv) the training curves of the RL agents in Fig. C.9.

Data availability

Data and code was shared in a repository indicated in the manuscript.

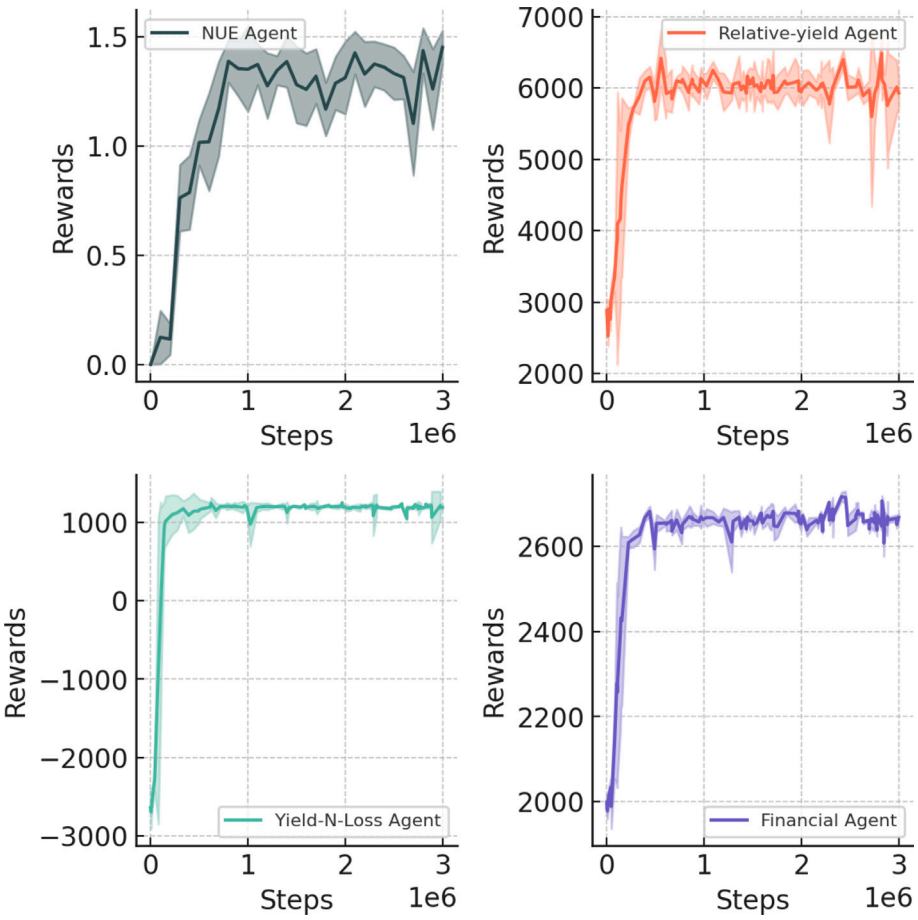

Fig. C.9. Training curves of all RL agents. The curves include all testing years and seeds. Each agent is plotted separately due to the different reward functions.

References

Abbas, S., Javed, M.T., Ali, Q., Azeem, M., Ali, S., et al., 2021. Nutrient deficiency stress and relation with plant growth and development. In: Engineering Tolerance in Crop Plants Against Abiotic Stress. CRC Press, pp. 239–262. http://dx.doi.org/10.1201/9781003160717-12.

Achiam, J., Held, D., Tamar, A., Abbeel, P., 2017. Constrained policy optimization. In: International Conference on Machine Learning. PMLR, pp. 22–31.

Altenbach, S., DuPont, F., Kothari, K., Chan, R., Johnson, E., Lieu, D., 2003. Temperature, water and fertilizer influence the timing of key events during grain development in a us spring wheat. J. Cereal Sci. 37, 9–20. http://dx.doi.org/10.1006/jcrs.2002.0483.

Arora, S., Doshi, P., 2021. A survey of inverse reinforcement learning: Challenges, methods and progress. Artificial Intelligence 297, 103500. http://dx.doi.org/10.1016/j.artint.2021.103500.

Baja, H., Kallenberg, M., Athanasiadis, I.N., 2025. To measure or not: A cost-sensitive, selective measuring environment for agricultural management decisions with reinforcement learning. http://dx.doi.org/10.48550/arXiv.2501.12823.

Berghuijs, H.N., Silva, J.V., Reidsma, P., de Wit, A.J., 2024. Expanding the wofost crop model to explore options for sustainable nitrogen management: A study for winter wheat in the Netherlands. Eur. J. Agron. 154, 127099. http://dx.doi.org/10.1016/j.eja.2024.127099.

Blackshaw, R.E., Molnar, L.J., Janzen, H.H., 2004. Nitrogen fertilizer timing and application method affect weed growth and competition with spring wheat. Weed Sci. 52, 614–622. http://dx.doi.org/10.1614/WS-03-104R.

Bohachevsky, I.O., Johnson, M.E., Stein, M.L., 1986. Generalized simulated annealing for function optimization. Technometrics 28, 209–217. http://dx.doi.org/10.2307/1269076.

Booth, S., Knox, W.B., Shah, J., Niekum, S., Stone, P., Allievi, A., 2023. The perils of trial-and-error reward design: Misdesign through overfitting and invalid task specifications. In: Proceedings of the 37th AAAI Conference on Artificial Intelligence. AAAI, pp. 5920–5929. http://dx.doi.org/10.1609/aaai.v37i5.25733.

Bouldin, D., Klausner, S., Reid, W., 1984. Use of Nitrogen from Manure. John Wiley & Sons, Ltd., pp. 221–245. http://dx.doi.org/10.2134/1990.nitrogenincropproduction.c15, chapter 15.

Camps-Valls, G., Fernández-Torres, M.Á., Cohrs, K.H., Höhl, A., Castelletti, A., Pacal, A., Robin, C., Martinuzzi, F., Papoutsis, I., Prapas, I., et al., 2025. Artificial intelligence for modeling and understanding extreme weather and climate events. Nat. Commun. 16, 1919. http://dx.doi.org/10.1038/s41467-025-56573-8.

Ceglar, A., Van der Wijngaart, R., De Wit, A., Lecerf, R., Boogaard, H., Seguini, L., Van den Berg, M., Toreti, A., Zampieri, M., Fumagalli, D., et al., 2019. Improving wofost model to simulate winter wheat phenology in Europe: Evaluation and effects on yield. Agricult. Sys. 168, 168–180. http://dx.doi.org/10.1016/j.agsy.2018.05.002.

Censor, Y., 1977. Pareto optimality in multiobjective problems. Appl. Math. Optim. 4, 41–59. http://dx.doi.org/10.1007/BF01442131.

Chen, X., Cui, Z., Fan, M., Vitousek, P., Zhao, M., Ma, W., Wang, Z., Zhang, W., Yan, X., Yang, J., et al., 2014. Producing more grain with lower environmental costs. Nature 514, 486–489. http://dx.doi.org/10.1038/nature13609.

Chen, Y., Lin, M., Yu, Z., Sun, W., Fu, W., He, L., 2025. Enhancing cotton irrigation with distributional actor–critic reinforcement learning. Agricult. Water. Manag. 307, 109194. http://dx.doi.org/10.1016/j.agwat.2024.109194.

Chen, Y., Yu, Z., Han, Z., Sun, W., He, L., 2023. A decision-making system for cotton irrigation based on reinforcement learning strategy. Agronomy 14, 11. http://dx.doi.org/10.3390/agronomy14010011.

CLO, 2022. Stikstofdepositie. pp. 1990–2022, https://www.clo.nl/indicatoren/nl0189-stikstofdepositie.

Cui, Z., Zhang, F., Chen, X., Dou, Z., Li, J., 2010. In-season nitrogen management strategy for winter wheat: Maximizing yields, minimizing environmental impact in an over-fertilization context. Field Crop. Res. 116, 140–146. http://dx.doi.org/10.1016/j.fcr.2009.12.004.

de Wit, A., 2023. The Python crop simulation environment.

De Wit, A., Boogaard, H., Fumagalli, D., Janssen, S., Knapen, R., van Kraalingen, D., Supit, I., van der Wijngaart, R., van Diepen, K., 2019. 25 years of the WOFOST cropping systems model. Agricult. Sys. 168, 154–167. http://dx.doi.org/10.1016/j.agsy.2018.06.018.

Eschmann, J., 2021. Reward function design in reinforcement learning. In: Reinforcement Learning Algorithms: Analysis and Applications. pp. 25–33. http://dx.doi.org/10.1007/978-3-030-41188-6_3.

EU Nitrogen Expert Panel, 2015. Nitrogen Use Efficiency (Nue) an Indicator for the Utilization of Nitrogen in Food Systems. Wageningen University, Alterra, Wageningen, Netherlands.

European Union, 1991. Council directive 91/676/eec of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources. Off. J. Eur. Communities L 375, 1–8, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31991L0676.

European Union, 2016. Directive (eu) 2016/2284 of the European parliament and of the council of 14 December 2016 on the reduction of national emissions of certain atmospheric pollutants, amending directive 2003/35/ec and repealing directive 2001/81/ec. Off. J. Eur. Communities L 344, 1–31, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L2284.

Faber, A., Jarosz, Z., Rutkowska, A., Jadczyszyn, T., 2021. Reduction of nitrogen losses in winter wheat grown on light soils. Agronomy 11. http://dx.doi.org/10.3390/agronomy11112337.

Fageria, N.K., Baligar, V.C., 2005. Enhancing nitrogen use efficiency in crop plants. Adv. Agron. 88, 97–185. http://dx.doi.org/10.1016/S0065-2113(05)88004-6.

FAO, 2019. The International Code of Conduct for the sustainable use and management of fertilizers. Food Agric. Organ. http://dx.doi.org/10.4060/CA5253EN.

Fetting, C., 2020. The European Green Deal. ESDN Report, December 2.

Fisher, M.L., 1981. The lagrangian relaxation method for solving integer programming problems. Manag. Sci. 27, 1–18. http://dx.doi.org/10.1287/mnsc.27.1.1.

Flach, B., Selten, M., 2021. Netherlands: Dutch Parliament Approves Law To Reduce Nitrogen Emissions. Global Agricultural Information Network Report NL2020-0069, Foreign Agricultural Service, US Department of Agriculture.

Fountas, S., Aggelopoulou, K., Gemtos, T.A., 2015. Precision agriculture: Crop management for improved productivity and reduced environmental impact or improved sustainability. Supply Chain Manag. Sustain. Food Netw. 4, 1–65. http://dx.doi.org/10.1002/9781118937495.ch2.

Gaso, D.V., de Wit, A., de Bruin, S., Puntel, L.A., Berger, A.G., Kooistra, L., 2023. Efficiency of assimilating leaf area index into a soybean model to assess within-field yield variability. Eur. J. Agron. 143, 126718. http://dx.doi.org/10.1016/j.eja.2022.126718.

Gautron, R., Maillard, O.A., Preux, P., Corbeels, M., Sabbadin, R., 2022a. Reinforcement learning for crop management support: Review, prospects and challenges. Comput. Electron. Agric. 200, 107182. http://dx.doi.org/10.1016/j.compag.2022.107182.

Gautron, R., Padrón, E.J., Preux, P., Bigot, J., Maillard, O.A., Emukpere, D., 2022b. gym-DSSAT: a crop model turned into a reinforcement learning environment. http://dx.doi.org/10.48550/arXiv.2207.03270, arXiv:2207.03270.

Giller, K.E., 2001. Nitrogen Fixation in Tropical Cropping Systems. Cabi.

Goldenits, G., Mallinger, K., Raubitzek, S., Neubauer, T., 2024. Current applications and potential future directions of reinforcement learning-based digital twins in agriculture. Smart Agric. Technol. 100512. http://dx.doi.org/10.1016/j.atech.2024.100512.

Groot, J., Verberne, E., 1991. Response of wheat to nitrogen fertilization, a data set to validate simulation models for nitrogen dynamics in crop and soil. In: Nitrogen Turnover in the Soil-Crop System: Modelling of Biological Transformations, Transport of Nitrogen and Nitrogen Use Efficiency. Proceedings of a Workshop held at the Institute for Soil Fertility Research, Haren, the Netherlands, 5–6 1990, Springer, pp. 349–383. http://dx.doi.org/10.1007/BF01051140.

Guertal, E., 2009. Slow-release nitrogen fertilizers in vegetable production: a review. HortTechnology 19, 16–19. http://dx.doi.org/10.21273/HORTTECH.19.1.16.

Hayes, C.F., Rădulescu, E., Källström, J., Macfarlane, M., Reymond, M., Verstraeten, T., Zintgraf, L.M., Dazeley, R., Heintz, F., et al., 2022. A practical guide to multi-objective reinforcement learning and planning. Auton. Agents Multi-Agent Syst. 36, 26. http://dx.doi.org/10.1007/s10458-022-09552-y.

Henaff, M., Raileanu, R., Jiang, M., Rocktäschel, T., 2023. Exploration via elliptical episodic bonuses. http://dx.doi.org/10.48550/arXiv.2210.05805.

Huang, B., Sun, W., Zhao, Y., Zhu, J., Yang, R., Zou, Z., Ding, F., Su, J., 2007. Temporal and spatial variability of soil organic matter and total nitrogen in an agricultural ecosystem as affected by farming practices. Geoderma 139, 336–345. http://dx.doi.org/10.1016/j.geoderma.2007.02.012.

Iizumi, T., Ramankutty, N., 2016. Changes in yield variability of major crops for 1981–2010 explained by climate change. Environ. Res. Lett. 11, 034003. http://dx.doi.org/10.1088/1748-9326/11/3/034003.

Ji, J., Zhang, B., Zhou, J., Pan, X., Huang, W., Sun, R., Geng, Y., Zhong, Y., Dai, J., Yang, Y., 2023. Safety-gymnasium: A unified safe reinforcement learning Benchmark. http://dx.doi.org/10.48550/arXiv.2310.12567, arXiv preprint arXiv:2310.12567.

Kallenberg, M.G., Overweg, H., Bree, R.van., Athanasiadis, I.N., 2023. Nitrogen management with reinforcement learning and crop growth models. Environ. Data Sci. 2, e34. http://dx.doi.org/10.1017/eds.2023.28.

Klages, S., Heidecke, C., Osterburg, B., Bailey, J., Calciu, I., Casey, C., Dalgaard, T., Frick, H., Glavan, M., D'Haene, K., Hofman, G., Leitão, I.A., Surdyk, N., Verloop, K., Velthof, G., 2020. Nitrogen surplus—a unified indicator for water pollution in europe? Water 12. http://dx.doi.org/10.3390/w12041197.

Kowalenko, C., Bittman, S., 2000. Within-season grass yield and nitrogen uptake, and soil nitrogen as affected by nitrogen applied at various rates and distributions in a high rainfall environment. Can. J. Plant Sci. 80, 287–301. http://dx.doi.org/10.4141/P98-139.

Lipper, L., Thornton, P., Campbell, B.M., Baedeker, T., Braimoh, A., Bwalya, M., Caron, P., Cattaneo, A., Garrity, D., Henry, K., et al., 2014. Climate-smart agriculture for food security. Nat. Clim. Chang. 4, 1068–1072. http://dx.doi.org/10.1038/nclimate2437.

Liu, Y., Halev, A., Liu, X., 2021. Policy learning with constraints in model-free reinforcement learning: A survey. In: The 30th International Joint Conference on Artificial Intelligence. ijcai, pp. 4508–4515. http://dx.doi.org/10.24963/ijcai.2021/614.

Luo, W., Li, H., Zhang, Z., Han, C., Lv, J., Guo, T., 2024. Sambo-rl: Shifts-aware model-based offline reinforcement learning. http://dx.doi.org/10.48550/arXiv.2408.12830, arXiv preprint arXiv:2408.12830.

Madondo, M., Azmat, M., Dipietro, K., Horesh, R., Jacobs, M., Bawa, A., Srinivasan, R., O'Donncha, F., 2023. A SWAT-based reinforcement learning framework for crop management. http://dx.doi.org/10.48550/arXiv.2302.04988, arXiv preprint arXiv:2302.04988.

Maillard, O.A., Mathieu, T., Basu, D., 2023. Farm-gym: A modular reinforcement learning platform for stochastic agronomic games. In: AIAFS 2023-Artificial Intelligence for Agriculture and Food Systems. p. 7.

Norton, R., Davidson, E., Roberts, T., 2015. Nitrogen Use Efficiency and Nutrient Performance Indicators. Global Partnership on Nutrient Management, p. 14.

Overweg, H., Berghuijs, H.N.C., Athanasiadis, I.N., 2021. CropGym: a reinforcement learning environment for crop management. http://dx.doi.org/10.48550/arXiv.2104.04326, CoRR abs/2104.04326.

Pathak, D., Agrawal, P., Efros, A.A., Darrell, T., 2017. Curiosity-driven exploration by self-supervised prediction. In: International Conference on Machine Learning. PMLR, pp. 2778–2787. http://dx.doi.org/10.48550/arXiv.1705.05363.

Pylianidis, C., Osinga, S., Athanasiadis, I.N., 2021. Introducing digital twins to agriculture. Comput. Electron. Agric. 184, 105942. http://dx.doi.org/10.1016/j.compag.2020.105942.

Raffin, A., Hill, A., Ernestus, M., Gleave, A., Kanervisto, A., Dormann, N., 2019. Stable baselines3. https://github.com/DLR-RM/stable-baselines3.

Ravensbergen, A.P.P., van Ittersum, M.K., Hijbeek, R., Kempenaar, C., Reidsma, P., 2024. Field monitoring reveals scope to reduce environmental impact of ware potato cultivation in the Netherlands without compromising yield. Agricult. Sys. 220, 104091. http://dx.doi.org/10.1016/j.agsy.2024.104091.

Raza, S., Farmaha, B.S., 2022. Contrasting corn yield responses to nitrogen fertilization in southeast coastal plain soils. Front. Environ. Sci. 10, 955142. http://dx.doi.org/10.3389/fenvs.2022.955142.

Roijers, D.M., Vamplew, P., Whiteson, S., Dazeley, R., 2013. A survey of multi-objective sequential decision-making. J. Artificial Intelligence Res. 48. http://dx.doi.org/10.1613/jair.3987.

Rosenthal, R.E., 1985. Concepts theory, and techniques principles of multiobjective optimization. Decis. Sci. 16, 133–152. http://dx.doi.org/10.1111/j.1540-5915.1985.tb01479.x.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal policy optimization algorithms. http://dx.doi.org/10.48550/arXiv.1707.06347, arXiv preprint arXiv:1707.06347.

Semenov, M.A., Barrow, E.M., Lars-Wg, A., 2002. A Stochastic Weather Generator for Use in Climate Impact Studies. User Man Herts UK, pp. 1–27.

Silva, J.V., Tenreiro, T.R., van Ittersum, M.K., Anten, N.P., Reidsma, P., 2018. Benchmarking nitrogen use efficiency of ware potato and winter wheat cropping systems in the Netherlands. In: Towards Zero Hunger: Partnerships for Impact. p. 1. http://dx.doi.org/10.13140/RG.2.2.35580.39049.

Silva, J.V., van Ittersum, M.K., ten Berge, H.F., Spätjens, L., Tenreiro, T.R., Anten, N.P., Reidsma, P., 2021. Agronomic analysis of nitrogen performance indicators in intensive arable cropping systems: An appraisal of big data from commercial farms. Field Crop. Res. 269, 108176. http://dx.doi.org/10.1016/j.fcr.2021.108176.

Song, X., Jiang, Y., Tu, S., Du, Y., Neyshabur, B., 2019. Observational overfitting in reinforcement learning. http://dx.doi.org/10.48550/arXiv.1912.02975, arXiv preprint arXiv:1912.02975.

Sutton, R.S., Barto, A.G., 2018. Reinforcement Learning: An Introduction. MIT Press.

Tan, Z.X., Lal, R., Wiebe, K.D., 2005. Global soil nutrient depletion and yield reduction. J. Sustain. Agric. 26, 123–146. http://dx.doi.org/10.1300/J064v26n01_10.

Tao, R., Zhao, P., Wu, J., Martin, N.F., Harrison, M.T., Ferreira, C., Kalantari, Z., Hovakimyan, N., 2022. Optimizing crop management with reinforcement learning and imitation learning. http://dx.doi.org/10.48550/arXiv.2209.09991, arXiv preprint arXiv:2209.09991.

Taylor, M.E., Stone, P., 2009. Transfer learning for reinforcement learning domains: A survey. J. Mach. Learn. Res. 10. http://dx.doi.org/10.5555/1577069.1755839.

Tessler, C., Mankowitz, D.J., Mannor, S., 2018. Reward constrained policy optimization. http://dx.doi.org/10.48550/arXiv.1805.11074, CoRR abs/1805.11074.

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P., 2017. Domain randomization for transferring deep neural networks from simulation to the real world. http://dx.doi.org/10.48550/arXiv.1703.06907, CoRR abs/1703.06907.

Towers, M., Kwiatkowski, A., Terry, J., Balis, J.U., De Cola, G., Deleu, T., Goulão, M., Kallinteris, A., Krimmel, M., KG, A., et al., 2024. Gymnasium: A standard interface for reinforcement learning environments. http://dx.doi.org/10.48550/arXiv.2407.17032, arXiv preprint arXiv:2407.17032.

Tubiello, F.N., Salvatore, M., Ferrara, A.F., House, J., Federici, S., Rossi, S., Biancalani, R., Condor Golec, R.D., Jacobs, H., Flammini, A., et al., 2015. The contribution of agriculture, forestry and other land use activities to global warming, 1990–2012. Global Change Biol. 21, 2655–2660. http://dx.doi.org/10.1111/gcb.12865.

Turchetta, M., Corinzia, L., Sussex, S., Burton, A., Herrera, J., Athanasiadis, I., Buhmann, J.M., Krause, A., 2022. Learning long-term crop management strategies with CyclesGym. Adv. Neural Inf. Process. Syst. 35, 11396–11409. http://dx.doi.org/10.5555/3600270.3601098.

van Bussel, L.G., Grassini, P., Van Wart, J., Wolf, J., Claessens, L., Yang, H., Boogaard, H., de Groot, H., Saito, K., Cassman, K.G., et al., 2015. From field to atlas: upscaling of location-specific yield gap estimates. Field Crop. Res. 177, 98–108. http://dx.doi.org/10.1016/j.fcr.2015.03.005.

Van der Pol, F., 1992. Soil mining: An unseen contributor to farm income in southern mali. Judge 65.

Van der Velde, M., Nisini, L., 2019. Performance of the MARS-crop yield forecasting system for the European Union: Assessing accuracy, in-season, and year-to-year improvements from 1993 to 2015. Agricult. Sys. 168, 203–212. http://dx.doi.org/10.1016/j.agsy.2018.06.009.

Van Diepen, C.v., Wolf, J.v., Van Keulen, H., Rappoldt, C., 1989. WOFOST: a simulation model of crop production. Soil Use Manag. 5, 16–24. http://dx.doi.org/10.1111/j.1475-2743.1989.tb00755.x.

Wageningen Economic Research, 2023a. agrimatie.nl. price development of fertilizer. https://www.agrimatie.nl/SectorResultaat.aspx?subpubID=2232&sectorID=2233&themaID=2263. (Accessed 19 January 2024).

Wageningen Economic Research, 2023b. agrimatie.nl. price development of seeds and grains. https://www.agrimatie.nl/ThemaResultaat.aspx?subpubID=2289&themaID=2263. (Accessed 19 January 2024).

Wang, Y., Li, C., Li, Y., Zhu, L., Liu, S., Yan, L., Feng, G., Gao, Q., 2020. Agronomic and environmental benefits of nutrient expert on maize and rice in Northeast China. Environ. Sci. Pollut. Res. 27. http://dx.doi.org/10.1007/s11356-020-09153-w.

Wierzbicki, A.P., 1980. The use of reference objectives in multiobjective optimization. In: Multiple Criteria Decision Making Theory and Application: Proceedings of the Third Conference Hagen/KÖnigswinter, West Germany, August (1979) 20–24. Springer, pp. 468–486. http://dx.doi.org/10.1007/978-3-642-48782-8_32.

Wu, J., Tao, R., Zhao, P., Martin, N.F., Hovakimyan, N., 2022. Optimizing nitrogen management with deep reinforcement learning and crop simulations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1712–1720. http://dx.doi.org/10.48550/arXiv.2204.10394.

Yang, Z., Hu, Y., Zhang, S., Raza, S., Wei, X., Zhao, X., 2022. The thresholds and management of irrigation and fertilization earning yields and water use efficiency in maize, wheat, and rice in China: A meta-analysis (1990–2020). Agronomy 12. http://dx.doi.org/10.3390/agronomy12030709.

Ye, C., Zheng, G., Tao, Y., Xu, Y., Chu, G., Xu, C., Chen, S., Liu, Y., Zhang, X., Wang, D., 2024. Effect of soil texture on soil nutrient status and rice nutrient absorption in paddy soils. Agronomy 14, 1339. http://dx.doi.org/10.3390/agronomy14061339.

Zhang, X., Davidson, E.A., Mauzerall, D.L., Searchinger, T.D., Dumas, P., Shen, Y., 2015. Managing nitrogen for sustainable development. Nature 528, 51–59. http://dx.doi.org/10.1038/nature15743.

Zhang, C., Vinyals, O., Munos, R., Bengio, S., 2018. A study on overfitting in deep reinforcement learning. http://dx.doi.org/10.48550/arXiv.1804.06893, arXiv preprint arXiv:1804.06893.

Zhou, J., Gu, B., Schlesinger, W.H., Ju, X., 2016. Significant accumulation of nitrate in Chinese semi-humid croplands. Sci. Rep. 6 (25088), http://dx.doi.org/10.1038/srep25088.

Zhou, G., Ke, L., Srinivasa, S., Gupta, A., Rajeswaran, A., Kumar, V., 2023. Real world offline reinforcement learning with realistic data source. In: 2023 IEEE International Conference on Robotics and Automation. ICRA, IEEE, pp. 7176–7183. http://dx.doi.org/10.48550/arXiv.2210.06479.