Environmental Data Science (2023), 2: e34, 1–12
doi:10.1017/eds.2023.28

CAMBRIDGE UNIVERSITY PRESS

APPLICATION PAPER

Nitrogen management with reinforcement learning and crop growth models

Michiel G.J. Kallenberg, Hiske Overweg, Ron van Bree and Ioannis N. Athanasiadis

Laboratory of Geo-information Science and Remote Sensing, Wageningen University & Research, Wageningen, The Netherlands
Corresponding author: Michiel G.J. Kallenberg; Email: michiel.kallenberg@wur.nl

Received: 28 February 2023; **Revised:** 06 June 2023; **Accepted:** 02 August 2023

Keywords: Crop growth models; nitrogen; reinforcement learning; smart farming; winter wheat

Abstract

The growing need for agricultural products and the challenges posed by environmental and economic factors have created a demand for enhanced agricultural systems management. Machine learning has increasingly been leveraged to tackle agricultural optimization problems, and in particular, reinforcement learning (RL), a subfield of machine learning, seems a promising tool for data-driven discovery of future farm management policies. In this work, we present the development of *CropGym*, a Gymnasium environment, where a reinforcement learning agent can learn crop management policies using a variety of process-based crop growth models. As a use case, we report on the discovery of strategies for nitrogen application in winter wheat. An RL agent is trained to decide weekly on applying a discrete amount of nitrogen fertilizer, with the aim of achieving a balance between maximizing yield and minimizing environmental impact. Results show that close to optimal strategies are learned, competitive with standard practices set by domain experts. In addition, we evaluate, as an out-of-distribution test, whether the obtained policies are resilient against a change in climate conditions. We find that, when rainfall is sufficient, the RL agent remains close to the optimal policy. With *CropGym,* we aim to facilitate collaboration between the RL and agronomy communities to address the challenges of future agricultural decision-making.

Impact Statement

This study presents *CropGym*, an open simulation environment to conduct reinforcement learning research for discovering adaptive, data-driven policies for farm management using a variety of process-based crop growth models. With a use case on nitrogen management, we demonstrate the potential of RL to learn sustainable policies that are competitive with standard practices set by domain experts.

1. Introduction

In recent years, smart farming technologies have been considered key enablers to reduce the usage of chemicals (fertilizers and plant protection products) and to reduce greenhouse gas emissions to enable reaching the Green Deal targets (Saiz-Rubio and Rovira-Más, 2020). A promising direction within smart farming technology research focuses on developing decision support systems (DSSs). These human–

This research article was awarded an Open Data and Open Materials badge for transparent practices. See the Data Availability Statement for details.

computer systems aim at providing farmers with a list of advice for supporting their business or organizational decision-making activities to optimize returns on inputs while preserving resources, within (environmental) constraints. With the evolution of agriculture into Agriculture 4.0, thanks to the employment of current technologies such as Internet of things, remote sensing, big data, and artificial intelligence, DSSs of various kinds have found their way into agriculture. Examples include, but are not limited to, applications for agricultural mission planning, climate change adaptation, food waste control, plant protection, and resource management of water and nutrients (Zhai et al., 2020).

The backbone of a DSS typically consists of a set of models that provide a representation of the environment and processes therein that are to be optimized. In particular, for resource management a substantial share of DSSs is based on process-based crop growth models (Graeff et al., 2012). These models mathematically describe the growth, development, and yield of a crop for given environmental conditions, such as type of soil, weather, and availability of water and nutrients. The scientific community offers numerous crop growth models with different levels of sophistication, limitations, and limits of applicability (Di Paola et al., 2016; Jones et al., 2017). Widely used frameworks are APSIM (Holzworth et al., 2014), DSSAT (Jones et al., 2003), and PCSE (de Wit, 2023), which contain models such as LINTUL-3 (Shibu et al., 2010) and WOFOST (de Wit et al., 2019).

Generally speaking, there are two major ways in which crop growth models are utilized in a DSS to derive crop management decisions (Gallardo et al., 2020): (1) Components in the model are exploited to provide estimates of crop yield-limiting factors, such as (future) deficiencies of nutrients and water, and (2) the model is employed as a specialized simulator to assess the impact of a set of (predefined) crop management practices. For both cases, it is not trivial to find the optimal set of actions, as decisions have to be made under uncertainty. For instance, driving factors for, for example, future nutrient uptake, such as future weather conditions, are uncertain at the time the model is asked for advice on fertilizer application.

Finding an optimized sequence of (crop management) decisions under uncertainty is a challenging task for which machine learning has increasingly been leveraged. In particular, reinforcement learning (RL), a subfield of machine learning, seems a relevant tool to tackle agricultural optimization problems (Binas et al., 2019). RL seeks to train intelligent agents in a trial-and-error fashion to take actions in an environment based on a reward signal. In RL, the environment is formally specified as a Markov decision process (MDP) $\{S, A, T, R\}$, with state space S, an available set of actions A, a transition function T, and a reward function R. In the context of, for example, crop management, S may consist of (virtual) measurements on the state of the crop, A may be dose of fertilizer to apply, T may be represented by a simulation step of a crop growth model, and R may be defined as the (projected) amount of yield.

Recently, a few research works have introduced RL for the management of agricultural systems. For instance, RL has been used for climate control in a greenhouse (Wang et al., 2020), planting, and pruning in a polyculture garden (Avigal et al., 2022), fertilizer (Overweg et al., 2021) and/or water management (Chen et al., 2021; Tao et al., 2022; Saikai et al., 2023), coverage path planning (Din et al., 2022), and crop planning (Turchetta et al., 2022) in open-field agriculture. A comprehensive overview of reinforcement learning for crop management support is given in Gautron et al. (2022b).

As is common practice in RL research in a pioneering stage, practically all mentioned works used simulated environments. Some of these environments have been made publicly available as software artifact. Examples that build on crop growth models include *CropGym* (Overweg et al., 2021), an interface to the Python Crop Simulation Environment (PCSE) (de Wit, 2023), *gym-DSSAT* (Gautron et al., 2022a), an integration of the DSSAT (Hoogenboom et al., 2019) crop models, *CropRL* (Ashcraft and Karra, 2021), a wrapper around the SIMPLE crop model (Zhao et al., 2019), *SWATGym* (Madondo et al., 2023), a wrapper around SWAT (Arnold et al., 2011), and *CyclesGym* (Turchetta et al., 2022), a wrapper around Cycles (Kemanian et al., 2022). The mentioned examples are implemented with the Gymnasium toolkit (Towers et al., 2023), which is a highly used framework for developing and comparing reinforcement learning algorithms. By providing standardized test beds, efforts like these are instrumental in further promoting and accelerating RL research for agricultural problems.

In this study, we present the development of *CropGym*, a Gymnasium environment, where a reinforcement learning agent can learn farm management policies using a variety of process-based crop

https://doi.org/10.1017/eds.2023.28 Published online by Cambridge University Press

growth models. In particular, we report on the discovery of strategies for nitrogen application in winter wheat and we evaluate the resiliency of the obtained policies against climate change. The focus on nitrogen is motivated by the fact that (in rain-fed winter wheat) nitrogen is a key driver for yield, yet if supplied in excessive amount it has a detrimental effect on the environment, including eutrophication of freshwater, groundwater contamination, tropospheric pollution related to emissions of nitrogen oxides and ammonia gas, and accumulation of nitrous oxide, a potent greenhouse gas (Zhang et al., 2015).

2. Methodology

2.1. CropGym

We developed *CropGym*, a Gymnasium environment, for farm management policies, such as fertilization and irrigation, using process-based crop growth models. *CropGym* is built around the Python Crop Simulation Environment (PCSE), a well-established open-source framework that includes implementations of a variety of crop simulation models. The software is characterized by a high level of customizability. Input parameters, such as crop characteristics, are easily configurable. For deriving driving variables, such as weather information, a broad selection of sources is available. Furthermore, dedicated routines facilitate the assimilation of observational data, such as field measurements. State parameters on crop growth and development, as well as carbon, water, and nutrient balances, are simulated and outputted at daily time steps. Farm management actions can be applied at the same resolution.

CropGym follows standard gym conventions and enables daily interactions between an RL agent and a crop model. The code is designed in a modular fashion and allows users to flexibly and easily create custom environments. Users can, for example, base action and reward functions on crop state variables, such as water stress, nitrogen uptake, and biomass. As a backbone, a variety of (components of) crop growth models can be selected or combined. *CropGym* is shipped with a set of preconfigured environments that allow for readily conducting RL research for farm management practices. The source code and documentation are available at https://www.cropgym.ai.

2.2. Use case

In this work, we present a use case on nitrogen management in rain-fed winter wheat. An agent was trained to decide weekly on applying a discrete amount of nitrogen fertilizer, with the goal of balancing the trade-off between yield and environmental impact.

In the following, we outline the components that comprise the environment of our use case.

State space S consists of the current state of the crop and a multidimensional weather observation, as parameterized with the variables listed in Table 1.

Action space A comprises three possible fertilizer application amounts, namely {0, 20, 40} kg/ha.

Table 1. *Crop growth and weather variables exposed in the state space S*

Variable	Meaning	Unit
DVS	Development stage	–
TGROWTH	Total biomass growth (above and below ground)	g/m2
LAI	Leaf area index	–
NUPTT	Total nitrogen uptake	–
TRAN	Transpiration	mm/day
TNSOIL	Total soil inorganic nitrogen	gN/m2
TRAIN	Total rainfall	mm
TRANRF	Transpiration reduction factor	–
WSO	Weight storage organs	g/m2
IRRAD	Incoming global radiation	J/m2/day
TMIN	Minimum temperature	°C
RAIN	Precipitation	cm/day

https://doi.org/10.1017/eds.2023.28 Published online by Cambridge University Press

Reward function R constitutes the balance between the gain in yield and the (environmental) costs associated with the application of nitrogen. *R* is formalized as follows:

$$r_t = (WSO_t^{\pi} - WSO_{t-1}^{\pi}) - (WSO_t^{0} - WSO_{t-1}^{0}) - \beta N_t, \tag{2.1}$$

with *t* the timestep, *WSO* the weight of the storage organ (g/m2), and *N* the amount of nitrogen (g/m2). The upper indices π and 0 refer to the agent's policy and a zero nitrogen policy, respectively. Parameter β determines the trade-off between increased yield and reduced environmental impact. Setting $\beta \approx 2.0$ corresponds to a reward that purely comprises economic profitability, since a kg of fertilizer is twice as expensive as a kg of wheat (Agri23a, 2023; Agri23b, 2023). In this work, we present results for $\beta = 10.0$ to emphasize the environmental costs.

Transitions are governed by the process-based crop model LINTUL-3 (light interception and utilization (Shibu et al., 2010)) and the weather sequence. The model parameters have been calibrated to simulate winter wheat in the Netherlands (Wiertsema, 2015; Berghuijs et al., 2023). Weather data were obtained from the PowerNASA database for three locations in the Netherlands and one in France for the years 1990 to 2022. An episode runs until the crop has reached maturity, which differs between episodes because of weather conditions.

An RL agent was trained with proximal policy optimization (PPO) (Schulman et al., 2017) as implemented in the Stable-Baselines3 library (Raffin et al., 2021). The environment was normalized with the VecNormalize environment wrapper, a normalized reward, and observation clipping set to 10. The discount factor γ was set to 1.0 as we aim to optimize the cumulative reward over the entire episode. To reduce redundancy among the input data, we aggregated the time-series data: The weather sequence, with size of 3x7 (i.e., features x days), was processed with an average pooling layer, yielding a feature vector of size of 3x1. The crop features, with size of 9x7, were shrunk to 9x1, by taking the last entry for each feature. Both resulting feature vectors were concatenated and subsequently flattened to obtain a feature vector of size of 12. The policy and the value network were a multilayer perceptron with two hidden layers, each of size of 128, and activation function tanh. Weights were shared between both networks. The training was done on the odd years from 1990 to 2022 (the even years were reserved for validation), with weather data from (52,5.5), (51.5,5), and (52.5,6.0) (°N,°E). The training ran for 400,000 timesteps using default hyperparameters. We selected PPO as our choice of RL algorithm due to its consistent high performance in RL research and its robust nature (Schulman et al., 2017). We also explored training a Deep Q-Network (DQN) (Mnih et al., 2013), which yielded similar results to those obtained (see Appendix A).

Two baseline agents were implemented as a reference for the RL agent:

The standard practice agent (SP) applies a fixed amount of nitrogen that is the same for all episodes. SP thereby reflects common practice, in which a predetermined amount of nitrogen is applied on three different dates during the season (Wiertsema, 2015). The static amount of nitrogen SP applied is determined by the optimization[1] on the training set.

The Ceres agent applies an episode-specific amount of nitrogen, that is optimized[1] for the episode it is evaluated on. Effectively, Ceres has access to the weather data of the entire season, which contrasts with

[1] For training the baseline agents such as Ceres and SP, we exploited a flaw in the nitrogen leaching component of LINTUL-3. In LINTUL-3, the nitrogen loss is computed as a fixed fraction of the amount of applied fertilizer (i.e., one minus fertilizer recovery fraction), regardless of timing and state dynamics, such as weather conditions. Any surplus of nitrogen is not leached, but remains available for uptake throughout the growing season. In principle, if we do not put constraints on the action space, we may apply all required nitrogen at once, right at the start of the season, thereby allowing for the elimination of the timing dimension of the problem. In this setting, optimization of the fertilization policy is reduced to finding the right amount of fertilizer, which can be resolved with a simple optimizer. Policies obtained by this strategy effectively mimic practices in which fertilizer is always applied in a timely manner, since the crop never has to wait before the applied fertilizer becomes available.

Note that we cannot employ the mentioned optimization regime when the action space is constrained. This premise is violated for the RL agent, as its action space is limited to a discrete amount of fertilizer, with a maximum of 40 kg/ha per action. This prevents the RL agent from applying all the fertilizer at once. Moreover, unlike the Ceres agent, the RL agent does not have access to future weather conditions and thus does not know the rewards of its actions in advance. As such, for the RL agent the timing dimension of the problem is preserved.

https://doi.org/10.1017/eds.2023.28 Published online by Cambridge University Press

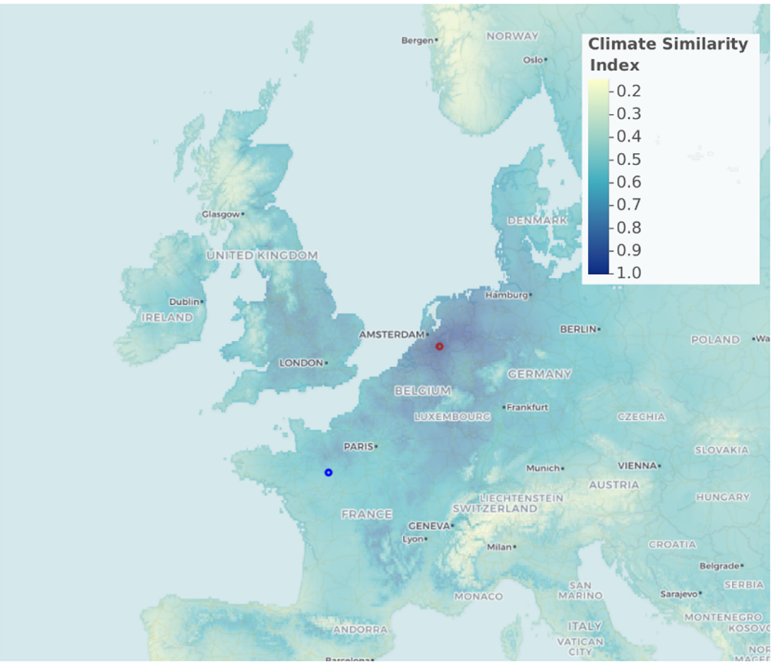

Figure 1. *Locations of the training (red) and out-of-distribution test (blue), with a CCAFS climate similarity index (Villegas et al., 2011) of 1.0 (reference) and 0.573, respectively.*

the RL agent that only has access to current and past weather data. Ceres can thus base its actions on future weather conditions and thereby reflects the upper bound of what any agent can achieve maximally.

We evaluated the performance of the implemented agents in the even years from 1990 to 2022 with weather data from (52,5.5) (°N,°E). As performance metrics, we computed the cumulative reward, the amount of nitrogen, and the yield, summarized as the median over the test years. For statistical analyses, we performed 15 runs initialized with different seeds and used bootstrapping to estimate the 95% confidence interval around the median.

As an out-of-distribution test, we evaluated the resiliency of the policy against a change in climate conditions. For that, we deployed both the trained RL and the baseline agents in a more southern climate. Practically, this was implemented by taking weather data from (48,0.0) (°N, °E), located in France, as opposed to (48,0.0) (°N, °E), located in the Netherlands, used during training (see Figure 1).

To reestablish the upper bound, Ceres was tuned to the weather data from the southern climate; SP and RL were not retrained. The robustness of the RL (and SP) agents was evaluated by assessing how close the agents' performance remains to the optimum, as determined by Ceres.

3. Results

A reinforcement learning agent (RL) was trained to find the optimal policy for applying nitrogen that balances yield increase and (environmental) costs. Two baseline agents were implemented for comparison: (1) The s*tandard practice* agent (SP) applies a fixed amount of nitrogen that does not differ between episodes and (2) the *Ceres* agent applies an episode-specific amount of nitrogen. The amount of nitrogen SP applied is determined by the optimization of the training set. Ceres, however, applies an amount of nitrogen that is optimized for the episode it is evaluated on. Ceres thereby reflects the upper bound of what any agent can achieve maximally.

https://doi.org/10.1017/eds.2023.28 Published online by Cambridge University Press

Table 2 reports the performance metrics for each of the three agents, summarized as the median over the test years, and its associated 95% confidence interval. The amount of nitrogen the RL agent applies, and the resulting yield and cumulative reward are close to the upper bound, as reflected by Ceres. Comparing the RL agent with the standard practice (SP), we see that RL applies more nitrogen, which results in a higher yield. The cumulative reward RL achieves is competitive with SP.

Figure 2 (left and middle) shows for each test year the reward obtained and amount of nitrogen applied by each of the three agents, as a function of the reward obtained and amount of nitrogen applied by Ceres. For most test years, RL is closer to Ceres than SP, in terms of both obtained cumulative reward and applied nitrogen. For two test years (2006 and 2010), the optimal amount of nitrogen, as determined by Ceres, is zero. In these years, which are characterized by a low amount of rainfall, the extra yield obtained by applying nitrogen does not outweigh the costs. RL (and SP) fail(s) to limit the nitrogen application, however, resulting in negative cumulative rewards. Figure 2 (right) shows for each test year the difference in yield between the RL agent and the SP agent, as a function of the difference in invested nitrogen. The diagonal shows the break-even line, for which the difference in reward is zero. Most test years, as well as the median, are above the break-even line, demonstrating that the RL agent's decision to apply a different amount of nitrogen is adequate.

Figure 3 shows the evolvement of the actions and rewards of the RL agent during the growing season, as summarized by the median over the test years. Typically, the RL agent waits until spring for its first actions. The median number of fertilization events is 7.0 (95% CI 6.0–8.0). The median length of an episode is 208 days (95% CI 205–211).

Table 2. *Cumulative reward, nitrogen, and yield (median and associated 95% CI)*

Agent	Cumulative reward		Nitrogen (kg/ha)		Yield (tonne/ha)	
Ceres	129.39	(73.34, 136.57)	183.0	(157.2, 211.1)	8.96	(8.41, 9.14)
SP	117.74	(67.47, 132.82)	170.7	(170.7, 170.7)	8.72	(8.13, 8.94)
RL	121.36	(67.96, 133.19)	180.0	(170.0, 200.0)	8.81	(8.24, 9.13)
$\Delta_{RL,SP}$	+3.33	(−1.94, 10.89) $p = 0.1057$	+9.3	(−0.70, 29.3) $p = 0.0398$	+0.13	(−0.01, 0.39) $p = 0.0290$

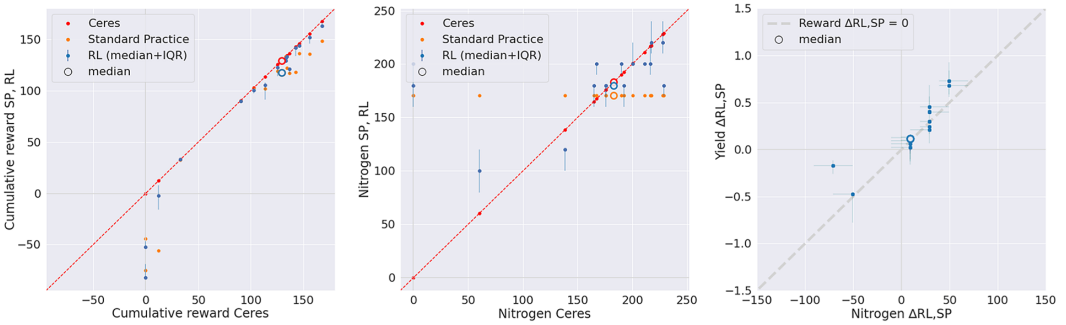

Figure 2. *(a): Cumulative reward obtained and (b): nitrogen applied by each of the three agents. Each dot depicts a test year (*n*=16). For most test years, RL is closer to Ceres than SP. (c): the difference in yield between the RL agent and the SP agent as a function of the difference in the amount of nitrogen applied. The dashed line indicates the break-even line, at which both agents achieve the same reward. Most test years are above the break-even line, demonstrating that the RL agent's choice of applying a different amount of nitrogen is adequate.*

https://doi.org/10.1017/eds.2023.28 Published online by Cambridge University Press

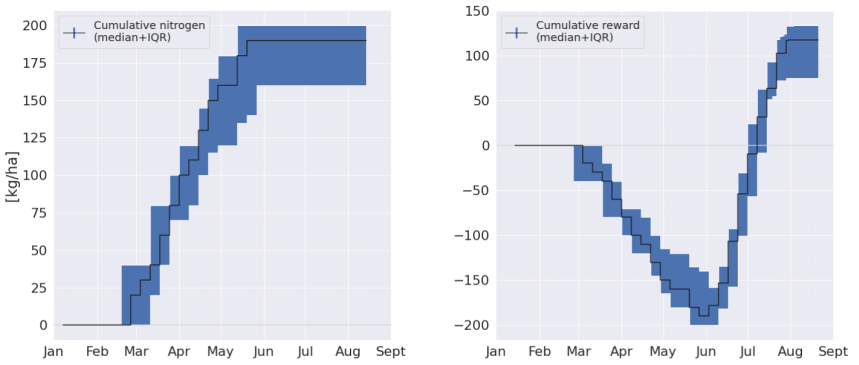

Figure 3. *Policy visualization of the RL agent: (a) cumulative reward obtained and (b) nitrogen applied. Typically, the RL agent waits until spring for its first actions.*

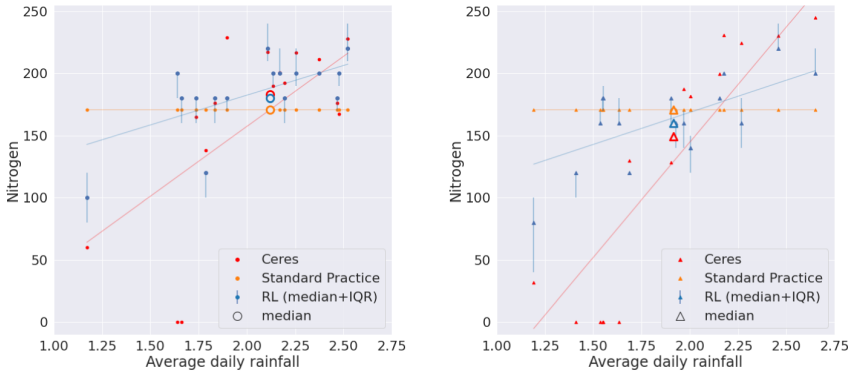

Figure 4. *Scatter plot with regression lines of the average daily rainfall and the total amount of nitrogen applied by all three agents for (a) the northern climate and (b) the southern climate. The optimal amount of nitrogen, as determined by Ceres, depends substantially on rainfall. Presumably, the RL agent has learned to adopt this general trend. In dry years, when lack of rainfall impairs yield and the optimal amount of nitrogen is (close to) zero, the RL agent does not limit its nitrogen application sufficiently, as it arguably sticks to the general trend. In years with sufficient rainfall, the RL agent acts in line with the optimal policy. This effect is seen in both the northern climate and the southern climate.*

The main driver for applying nitrogen is rainfall. Pearson's correlation coefficient between the total amount of rainfall during the growing season and the total amount of nitrogen applied is 0.69 (95% CI 0.58–0.76) for Ceres and 0.63 (95% CI 0.12–0.82) for RL (see Figure 4).

3.1. Climate resilience

To assess the robustness of the learned policy against changing climate conditions, we deployed both the trained RL and the baseline agents in a more southern climate. With a climate similarity index of 0.573, as determined with the CCAFS method (Villegas et al., 2011), using average temperature and precipitation as weather variables, the southern climate differs substantially from the northern climate. The southern climate is characterized by a higher average temperature (10.4 °C vs

https://doi.org/10.1017/eds.2023.28 Published online by Cambridge University Press

Table 3. *Out-of-distribution results: cumulative reward, nitrogen, and yield (median and 95% CI) in southern climate*

Agent	Cumulative reward		Nitrogen (kg/ha)		Yield (tonne/ha)	
Ceres	95.15	(0.0, 157.67)	149.6	(0.0, 202.99)	8.60	(4.69, 9.60)
SP	89.43	(−67.02, 140.23)	170.7	(170.7, 170.7)	8.50	(5.80, 9.13)
RL	78.17	(−49.92, 142.65)	160.0	(140.0, 180.0)	8.45	(5.80, 9.13)
$\Delta_{RL,SP}$	+4.52	(−4.85, 16.79) $p = 0.1792$	−10.70	(−30.7, 9.3) $p = 0.7244$	−0.04	(−0.17, 0.13) $p = 0.6459$

9.8 °C), less amount of average daily rainfall (1.92 mm vs 2.12 mm), and a shorter growing season (200 days vs 208 days).

Table 3 reports the performance metrics for each of the three agents deployed in the southern climate. The maximally achievable cumulative reward and yield, as represented by Ceres, are lower than what is obtained in the northern climate. In six years, namely 1990, 1992, 1996, 2004, 2006, and 2010, yield is (partially) limited by a low amount of rainfall, resulting in low cumulative rewards. In these dry years, the performance of the RL agent is suboptimal, as it does not limit its nitrogen application sufficiently. Yet, the cumulative reward RL achieves is competitive with SP. In years with sufficient rainfall, the RL agent remains close to the optimal policy, as is illustrated in Figure 4, just as we saw for the northern climate.

4. Discussion

We presented *CropGym*, a Gymnasium environment, to study policies for farm management, such as fertilization and irrigation, using process-based crop growth models. We developed a use case on nitrogen fertilization in rain-fed winter wheat. A reinforcement learning agent was trained to find the optimal timings and amounts for applying nitrogen that balance yield and environmental impact. The agent was found to learn close to optimal strategies, competitive with standard practices set by domain experts.

As an out-of-distribution test, we evaluated whether the obtained policies were resilient against a change in climate conditions, with sound results. Yet, in years where yield is limited by a shortage of rainfall, the performance of the RL agent was suboptimal. The adoption of more dry weather data in the training through, for example, fine-tuning approaches, may improve these results. Other examples of out-of-distribution tests with practical impact include variations in soil characteristics, such as organic matter content.

Clearly, as is common in RL research, our experiments are done in silico, and it is an open question to what extent our results transfer into the real world. (Crop growth) models are by definition simplifications of reality, and thus, policies derived from these models are inherently subject to a simulation-to-reality gap. Narrowing this gap can be achieved by employing an ensemble of different crop growth models (Wallach et al., 2018). *CropGym* supports such a strategy by offering implementations of a variety of process-based crop growth models.

To further bridge the gap between simulation and reality, digital twin technology could be exploited (Pylianidis et al., 2021). A variety of sensors can be employed to synchronize digital representations of crops with their physical counterparts (Jin et al., 2018; Jindo et al., 2023). Yet, the acquisition of sensor data may come with high (monetary) costs. In this context, *CropGym* could be utilized to train agents that are able to determine when and to what extent the environment should be measured (Bellinger et al., 2021). In such a training, the agent chooses between either relying on the simulated state of the crop or paying the cost to measure the true state and update the crop growth model accordingly.

https://doi.org/10.1017/eds.2023.28 Published online by Cambridge University Press

In this work, we incentivize the RL agent to generate environmentally friendly policies by negotiating the environmental costs of nitrogen application in the reward function. An alternative approach would be to set hard constraints on the total amount of nitrogen applied. Such could be achieved by building on the works in the domain of (safety)-constrained RL (Liu et al., 2021), supported by, for example, OpenAI's dedicated Safety Gym benchmark suite (Ray et al., 2019). Another constraint that could be considered is the number of fertilization events.

As an open simulation environment, *CropGym* can be used to discover adaptive, data-driven policies that perform well across a range of plausible scenarios for the future. With *CropGym,* we aim to facilitate a joint research effort from the RL and agronomy communities to meet the challenges of future agricultural decision-making and to further match farmers' decision-making processes.

Acknowledgments. The authors would like to thank Allard de Wit and Herman Berghuijs for the discussion and revision of the employed crop growth model and Lotte Woittiez for the constructive discussions during various stages of the research.

Author contribution. M.G.J.K., H.O., and I.N.A. conceptualized the study; M.G.J.K. and I.N.A. designed methodology; M.G.J.K., H.O., and R.v.B. provided software; M.G.J.K. visualized the data; M.G.J.K. and I.N.A. validated the data; M.G.J.K. wrote the original draft; H.O., R.v.B., and I.N.A. wrote, reviewed, and edited the article; and all authors approved the final submitted draft.

Competing interest. The authors declare none.

Data availability statement. Data and replication code can be found at https://www.cropgym.ai.

Ethics statement. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

Funding statement. This work has been partially supported by the European Union Horizon 2020 Research and Innovation program (Project Code: 101070496, Smart Droplets) and the Wageningen University and Research Investment Programme "Digital Twins."

References

Agri23a (2023) agrimatie.nl. price development of seeds and grains. Available at https://agrimatie.nl/SectorResultaat.aspx?subpubID=2232&sectorID=2233&themaID=2263 (accessed 19 January 2023).

Agri23b (2023) agrimatie.nl. price development of fertilizer. Available at https://agrimatie.nl/ThemaResultaat.aspx?subpubID=2289&themaID=2263 (accessed 19 January 2023).

Arnold JG, **Kiniry JR**, **Srinivasan R**, **Williams JR**, **Haney EB and Neitsch SL** (2011) Soil and Water Assessment Tool Input/Output File Documentation Version 2009. Technical report, Texas Water Resources Institute.

Ashcraft C and Karra K (2021) Machine learning aided crop yield optimization. Preprint, arXiv:2111.00963.

Avigal Y, **Wong W**, **Presten M**, **Theis M**, **Aeron S**, **Deza A**, **Sharma S**, **Parikh R**, **Oehme S**, **Carpin S**, **Viers JH**, **Vougioukas S and Goldberg KY** (2022) Simulating polyculture farming to learn automation policies for plant diversity and precision irrigation. *IEEE Transactions on Automation Science and Engineering 19*(3), 1352–1364. https://doi.org/10.1109/TASE.2021.3138995

Bellinger C, **Drozdyuk A**, **Crowley M and Tamblyn I** (2021) Scientific discovery and the cost of measurement - balancing information and cost in reinforcement learning. CoRR, abs/2112.07535. Available at arXiv preprint arXiv:2112.07535.

Berghuijs HNC, **Silva JV**, **Rijk HCA**, **van Ittersum MK**, **van Evert FK and Reidsma P** (2023) Catching-up with genetic progress: Simulation of potential production for modern wheat cultivars in the Netherlands. *Field Crops Research 296*, 108891. https://doi.org/10.1016/j.fcr.2023.108891; Available at https://www.sciencedirect.com/science/article/pii/S0378429023000849.

Binas J, **Luginbuehl L and Bengio Y** (2019) Reinforcement learning for sustainable agriculture. ICML 2019 Workshop Climate Change: How Can AI Help.

Chen M, **Cui Y**, **Wang X**, **Xie H**, **Liu F**, **Luo T**, **Zheng S and Luo Y** (2021) A reinforcement learning approach to irrigation decision-making for rice using weather forecasts. *Agricultural Water Management 250*, 106838. https://doi.org/10.1016/j.agwat.2021.106838; Available at https://www.sciencedirect.com/science/article/pii/S0378377421001037.

de Wit A (2023) The Python Crop Simulation Environment. Available at https://pcse.readthedocs.io/en/stable/.

de Wit A, **Boogaard H**, **Fumagalli D**, **Janssen S**, **Knapen R**, **van Kraalingen D**, **Supit I**, **van der Wijngaart R and van Diepen K** (2019) 25 years of the wofost cropping systems model. *Agricultural Systems 168*, 154–167. https://doi.org/10.1016/j.agsy.2018.06.018; Available at https://www.sciencedirect.com/science/article/pii/S0308521X17310107.

Di Paola A, **Valentini R and Santini M** (2016) An overview of available crop growth and yield models for studies and assessments in agriculture. *Journal of the Science of Food and Agriculture 96*(3), 709–714. https://doi.org/10.1002/jsfa.7359; Available at https://onlinelibrary.wiley.com/doi/abs/10.1002/jsfa.7359.

Din A, Ismail MY, Shah B, Babar M, Ali F and Baig SU (2022) A deep reinforcement learning-based multi-agent area coverage control for smart agriculture. *Computers and Electrical Engineering 101*, 108089. https://doi.org/10.1016/j.compeleceng.2022.108089; Available at https://www.sciencedirect.com/science/article/pii/S0045790622003445.

Gallardo M, Elia A and Thompson RB (2020) Decision support systems and models for aiding irrigation and nutrient management of vegetable crops. *Agricultural Water Management 240*, 106209. https://doi.org/10.1016/j.agwat.2020.106209; Available at https://www.sciencedirect.com/science/article/pii/S0378377420303267.

Gautron R, Gonzalez EJP, Preux P, Bigot J, Maillard O-A and Emukpere D (2022a) gym-DSSAT: A Crop Model Turned into a Reinforcement Learning Environment. PhD thesis, Inria Lille.

Gautron R, Maillard O-A, Preux P, Corbeels M and Sabbadin R (2022b) Reinforcement learning for crop management support: Review, prospects and challenges. *Computers and Electronics in Agriculture 200*, 107182. https://doi.org/10.1016/j.compag.2022.107182; Available at https://www.sciencedirect.com/science/article/pii/S0168169922004999.

Graeff S, Link J, Binder J and Claupein W (2012) Crop models as decision support systems in crop production. *Crop Production Technologies* 3–28. https://doi.org/10.5772/28976.

Holzworth DP, Huth NI, deVoil PG, Zurcher EJ, Herrmann NI, McLean G, Chenu K, van Oosterom EJ, Snow V, Murphy C, Moore AD, Brown H, Whish JPM, Verrall S, Fainges J, Bell LW, Peake AS, Poulton PL, Hochman Z, Thorburn PJ, Gaydon DS, Dalgliesh NP, Rodriguez D, Cox H, Chapman S, Doherty A, Teixeira E, Sharp J, Cichota R, Vogeler I, Li FY, Wang E, Hammer GL, Robertson MJ, Dimes JP, Whitbread AM, Hunt J, van Rees H, McClelland T, Carberry PS, Hargreaves JNG, MacLeod N, McDonald C, Harsdorf J, Wedgwood S and Keating BA (2014) Apsim – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software 62*, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009; Available at https://www.sciencedirect.com/science/article/pii/S1364815214002102.

Hoogenboom G, Porter CH, Boote KJ, Shelia V, Wilkens PW, Singh U, White JW, Asseng S, Lizaso JI, Patricia Moreno L, et al. (2019) The DSSAT crop modeling ecosystem. In *Advances in Crop Modelling for a Sustainable Agriculture*. Burleigh Dodds Science Publishing, pp. 173–216.

Jin X, Kumar L, Li Z, Feng H, Xu X, Yang G and Wang J (2018) A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy 92*, 141–152. https://doi.org/10.1016/j.eja.2017.11.002; Available at https://www.sciencedirect.com/science/article/pii/S1161030117301685.

Jindo K, Kozan O and de Wit A (2023) Data assimilation of remote sensing data into a crop growth model. In *Precision Agriculture: Modelling*. Springer, pp. 185–197.

Jones JW, Antle JM, Basso B, Boote KJ, Conant RT, Foster I, Godfray HCJ, Herrero M, Howitt RE, Janssen S, Keating BA, Munoz-Carpena R, Porter CH, Rosenzweig C and Wheeler TR (2017) Brief history of agricultural systems modeling. *Agricultural Systems 155*, 240–254. https://doi.org/10.1016/j.agsy.2016.05.014; Available at https://www.sciencedirect.com/science/article/pii/S0308521X16301585.

Jones JW, Hoogenboom G, Porter CH, Boote KJ, Batchelor WD, Hunt LA, Wilkens PW, Singh U, Gijsman AJ and Ritchie JT (2003) The DSSAT cropping system model. *European Journal of Agronomy 18*(3), 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7; Available at https://www.sciencedirect.com/science/article/pii/S1161030102001077.

Kemanian AR, White CM, Shi Y, Stockle CO and Leonard L (2022) Cycles: Agroecosystems model. Available at https://plantscience.psu.edu/research/labs/kemanian/models-and-tools/cycles.

Liu Y, Halev A and Liu X (2021) Policy learning with constraints in model-free reinforcement learning: A survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI-21)*, pp 4508–4515.

Madondo M, Azmat M, DiPietro K, Horesh R, Jacobs M, Bawa A, Srinivasan R and O'Donncha F (2023) A swat-based reinforcement learning framework for crop management. In *AAAI Conference on Artificial Intelligence*.

Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D and Riedmiller MA (2013) Playing atari with deep reinforcement learning. CoRR, abs/1312.5602; Available at http://arxiv.org/abs/1312.5602.

Overweg H, Berghuijs HNC and Athanasiadis IN (2021) Cropgym: A reinforcement learning environment for crop management. In *ICLR Workshop Modeling Oceans and Climate Change*.

Pylianidis C, Osinga S and Athanasiadis I (2021) Introducing digital twins to agriculture. *Computers and Electronics in Agriculture 184*, 105942. https://doi.org/10.1016/j.compag.2020.105942.

Raffin A, Hill A, Gleave A, Kanervisto A, Ernestus M and Dormann N (2021) Stable-baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research 22*(268), 1–8; Available at http://jmlr.org/papers/v22/20-1364.html.

Ray A, Achiam J and Amodei D (2019) Benchmarking safe exploration in deep reinforcement learning. Preprint, arXiv: 1910.01708.

Saikai Y, Peake A and Chenu K (2023) Deep reinforcement learning for irrigation scheduling using high-dimensional sensor feedback. Available at https://arxiv.org/abs/2301.00899.

Saiz-Rubio V and Rovira-Más F (2020) From smart farming towards agriculture 5.0: A review on crop data management. *Agronomy 10*(2). https://doi.org/10.3390/agronomy10020207; Available at https://www.mdpi.com/2073-4395/10/2/207.

Schulman J, Wolski F, Dhariwal P, Radford A and Klimov O (2017) Proximal policy optimization algorithms. CoRR, abs/1707.06347; Available at http://arxiv.org/abs/1707.06347.

Shibu ME, Leffelaar PA, Van Keulen H and Aggarwal P (2010) Lintul3, a simulation model for nitrogen-limited situations: Application to rice. *European Journal of Agronomy 32*, 255–271.

https://doi.org/10.1017/eds.2023.28 Published online by Cambridge University Press

Tao R, **Zhao P**, **Wu J**, **Martin NF**, **Harrison MT**, **Ferreira C**, **Kalantari Z and Hovakimyan N** (2022) Optimizing crop management with reinforcement learning and imitation learning. Preprint, arXiv:2209.09991.

Towers M, **Terry JK**, **Kwiatkowski A**, **Balis JU**, **Cola G**, **Deleu T**, **Goulão M**, **Kallinteris A**, **Arjun KG**, **Krimmel M**, **Perez-Vicente R**, **Pierré A**, **Schulhoff S**, **Tai JJ**, **Shen AJ and Younis OG** (2023) Gymnasium (v0.29.1). Zenodo. https://doi.org/10.5281/zenodo.8269265.

Turchetta M, **Corinzia L**, **Sussex S**, **Burton A**, **Herrera J**, **Athanasiadis IN**, **Buhmann JM and Krause A** (2022) Learning long-term crop management strategies with cyclesgym. In *Advances in Neural Information Processing Systems 35 (NeurIPS 2022)*.

Villegas JR, **Lau C**, **Köhler A-K**, **Jarvis A**, **Arnell NP**, **Osborne TM and Hooker J** (2011) Climate Analogues: Finding Tomorrow's Agriculture Today. CCAFS Working Paper.

Wallach D, **Martre P**, **Liu B**, **Asseng S**, **Ewert F**, **Thorburn PJ**, **van Ittersum M**, **Aggarwal PK**, **Ahmed M**, **Basso B**, **Biernath C**, **Cammarano D**, **Challinor AJ**, **De Sanctis G**, **Dumont B**, **Rezaei EE**, **Fereres E**, **Fitzgerald GJ**, **Gao Y**, **Garcia-Vila M**, **Gayler S**, **Girousse C**, **Hoogenboom G**, **Horan H**, **Izaurralde RC**, **Jones CD**, **Kassie BT**, **Kersebaum KC**, **Klein C**, **Koehler A-K**, **Maiorano A**, **Minoli S**, **Müller C**, **Kumar SN**, **Nendel C**, **O'Leary GJ**, **Palosuo T**, **Priesack E**, **Ripoche D**, **Rötter RP**, **Semenov MA**, **Stöckle C**, **Stratonovitch P**, **Streck T**, **Supit I**, **Tao F**, **Wolf J and Zhang Z** (2018) Multimodel ensembles improve predictions of crop–environment–management interactions. *Global Change Biology 24*(11), 5072–5083. https://doi.org/10.1111/gcb.14411; Available at https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.14411.

Wang L, **He X and Luo D** (2020) Deep reinforcement learning for greenhouse climate control. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*. IEEE, pp. 474–480.

Wiertsema W (2015) Obtaining Winter Wheat Parameters for Lintul from a Field Experiment. Master's thesis, Wageningen University, The Netherlands.

Zhai Z, **Martínez JF**, **Beltran V and Martínez NL** (2020) Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, *170*, 105256. https://doi.org/10.1016/j.compag.2020.105256; Available at https://www.sciencedirect.com/science/article/pii/S0168169919316497.

Zhang X, **Davidson EA**, **Mauzerall DL**, **Searchinger TD**, **Dumas P and Shen Y** (2015) Managing nitrogen for sustainable development. *Nature 528*(7580), 51–59.

Zhao C, **Liu B**, **Xiao L**, **Hoogenboom G**, **Boote KJ**, **Kassie BT**, **Pavan W**, **Shelia V**, **Kim KS**, **Hernandez-Ochoa IM**, et al. (2019) A simple crop model. *European Journal of Agronomy 104*, 97–106.

A. Appendix. Results of DQN

In addition to training with PPO, we explored training a Deep Q-Network (DQN) (Mnih et al., 2013). Configurations were kept the same as with PPO. We used the default settings of the (hyper)parameters, as set by Stable Baselines, except for (1) the number of hidden units, which was set to 128x128, (2) the activation function, which was set to tanh , and (3) *exploration_final_eps*, which was set to 0.01. The training ran for 400,000 timesteps.

Below, we report the key results, aggregated over five runs with different random seeds. Figure 5 demonstrates that for each test year (a) the cumulative reward, (b) the amount of fertilizer, and (c) the yield obtained by the DQN agent closely resemble those of the PPO agent. Table 4 shows that, similar to RL_{PPO}, also RL_{DQN} achieves results competitive with SP.

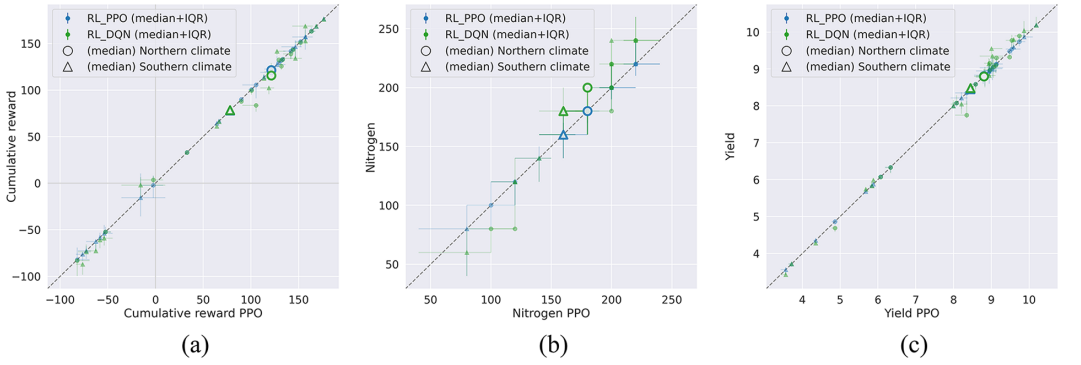

Figure 5. *Scatter plot of PPO and DQN agent for (a): cumulative reward obtained, (b): nitrogen applied, and (c): yield obtained. Each point depicts a test year (*n = *32).*

https://doi.org/10.1017/eds.2023.28 Published online by Cambridge University Press

Table 4. *Cumulative reward, nitrogen, and yield (median and associated 95% CI) for DQN*

Agent	Cumulative reward		Nitrogen (kg/ha)		Yield (tonne/ha)	
Northern climate						
RL_{DQN}	115.60	(59.73, 132.75)	200.0	(170.0, 210.0)	8.80	(8.18, 9.21)
$\Delta_{RL_{DQN},SP}$	+0.38	(−5.05, 9.87)	+29.30	(−0.70, 39.30)	+0.16	(−0.04, 0.41)
		$p = 0.4822$		$p = 0.0355$		$p = 0.0360$
Southern climate						
RL_{DQN}	78.46	(−54.88, 144.71)	180.0	(150.0, 190.0)	8.48	(5.83, 9.33)
$\Delta_{RL_{DQN},SP}$	+1.34	(−6.85, 15.83)	+9.30	(−20.70, 19.30)	+0.02	(−0.18, 0.15)
		$p = 0.4043$		$p = 0.4734$		$p = 0.4726$

Cite this article: Kallenberg MGJ, Overweg H, van Bree R and Athanasiadis IN (2023). Nitrogen management with reinforcement learning and crop growth models. *Environmental Data Science*, 2: e34. doi:10.1017/eds.2023.28