Reinforcement Learning for Sustainable Crop Management under Uncertainty

Vanessa Lopez-Corona[1], Yayong Li[1], Qiaomin Chen[2], Scott Chapman[2], Nan Ye[1]

[1]School of Mathematics and Physics, The University of Queensland
[2]School of Agriculture and Food Sustainability, The University of Queensland

OVERVIEW

Nitrogen management in rain-fed wheat, such as in Australia, is a sequential decision problem under uncertain weather. Reinforcement learning (RL) can learn nitrogen policies from simulated seasons in crop simulators, with the aim of maintaining yield while improving profitability in ways growers can apply.

Motivation

- Nitrogen is a major input cost: over-application reduces profit and causes environmental losses, while under-application limits yield.
- Fixed schedules and expert rules do not adapt to seasonal variation.
- Learning adaptive strategies directly on farms is impractical, as it would require many seasons of trial and error.
- Many approaches rely on destructive crop measurements, which are not always available when decisions are made.

Contributions

- An RL framework for nitrogen management in WOFOST, extended with Australian price ratios and a fixed cost per application.
- A sensitivity analysis showing learned policies perform similarly across nitrogen-to-wheat price ratios.
- A non-destructive policy using only crop development stage (DVS): a DVS rule matches standard practice, while RL policies come close in training years but not in unseen years.

Methodology

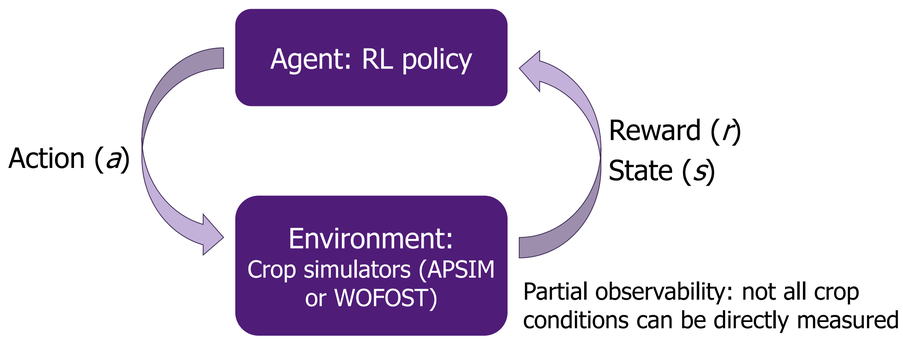


Figure 1: At each weekly step, the agent selects a nitrogen rate and receives the updated crop state and a reward. Over many seasons, it learns which actions lead to better outcomes.

Winter wheat in the Netherlands was simulated using the WOFOST crop model with the SNOMIN soil nitrogen module, through the PCSE-Gym environment. The environment setup is adapted from Kallenberg et al. (2023):

Observation: crop variables (DVS, LAI, WSO, N uptake, NO3, NH4, N-loss, soil N) and weather (IRRAD, TMIN, RAIN). Sensitivity analysis policies use all crop variables; the non-destructive policy uses DVS as its only crop variable.

Action: 0, 30, 60 or 90 kg N/ha, selected weekly

Agent: Proximal Policy Optimization (PPO, Stable-Baselines3)

Reward: extra yield over a no-fertiliser baseline, minus nitrogen and application costs

This work extends the setup with Australian price ratios and a fixed cost per application (machinery and labour), so more frequent applications cost more. Policies were trained at two nitrogen price ratios (3.43 and 6) and evaluated at both ratios over 5 years (3 for training, 2 for testing).

Non-Destructive Observations Policy

An RL agent was trained with development stage (DVS) as its only crop variable, which can be aligned with standard practice. Standard practice applies three doses of 60 kg N/ha on fixed dates. For comparison, a DVS threshold policy applies the same doses at the first weekly step where DVS reaches 0.07, 0.19 and 0.48. It was trained with 10 random seeds on 3 years and tested on 2; results are shown for the 3 policies with the highest training reward.

Year	Standard practice	DVS threshold	DVS-only RL (top 3)
1999 (train)	3100	3099	3056 ± 66
2000 (test)	2461	2460	1771 ± 392
2001 (train)	2468	2469	2349 ± 36
2002 (test)	2237	2239	1925 ± 348
2003 (train)	2181	2182	2110 ± 63

Table 1: Reward per year; RL values are mean ± SD of the top 3 seeds.

- The DVS threshold policy matched standard practice in every year, showing that a comparable policy based only on DVS exists.
- RL policies came close to standard practice in training years but performed worse and varied more in test years, so such a policy was not yet reliably learned.

In stable environments such as the Netherlands, standard practice is difficult to improve upon; the main challenge is learning policies that generalise to unseen years.

Sensitivity Analysis

Nitrogen and wheat prices both change over time. Their ratio (nitrogen price ÷ wheat price) sets how much extra grain each kg of nitrogen must produce to pay for itself. A ratio of 3.43 reflects current prices (1.2 ÷ 0.35 AUD/kg); a ratio of 6 represents a rise in fertiliser cost, a major concern for growers.

This tests how sensitive learned policies are to the price ratio. A policy that remains near-optimal across prices is desirable, as it would not need retraining when prices change, although policies can also be retrained for specific prices.

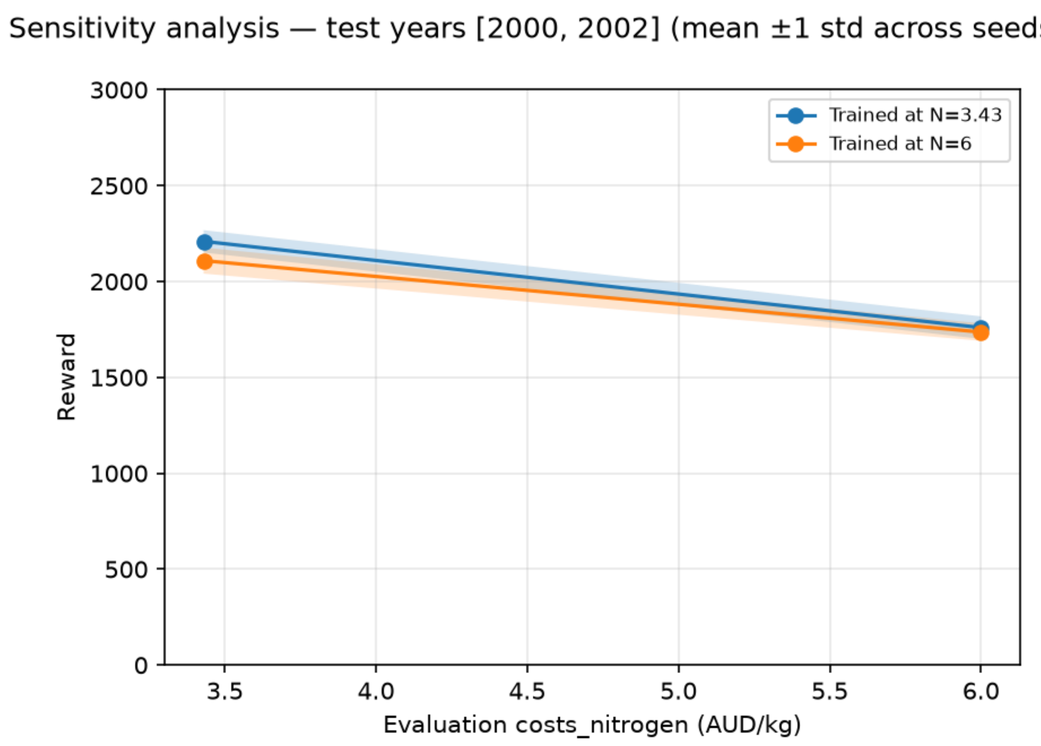


Figure 2: Reward in the test years (2000, 2002) for policies trained at price ratios of 3.43 (blue) and 6 (orange), evaluated at both prices. Mean ± 1 SD across 3 seeds.

Policies performed similarly under both price ratios, suggesting that a single policy can perform well across changes in the nitrogen-to-wheat price ratio.

Next Steps

- Improve generalisation to unseen years by training on more seasons.
- Extend to Australian conditions, where higher seasonal variability may give adaptive policies more room to improve.
- Compare WOFOST and APSIM to separate site effects from simulator effects.
- Extend to multi-year simulations that capture longer-term soil and climate dynamics.

References

Kallenberg, M. G. J., Overweg, H., van Bree, R., & Athanasiadis, I. N. (2023). Nitrogen management with reinforcement learning and crop growth models. Environmental Data Science, 2, e34. doi:10.1017/eds.2023.28

The support of the Grains Research and Development Corporation (GRDC) provided through the Analytics for the Australian Grains Industry (AAGI) project is acknowledged.