Grenze International Journal of Engineering and Technology, June Issue

Intelligent Fertilizer Optimization in Precision Agriculture using Deep Reinforcement Learning for Sustainable Crop Production

Suyash Kumar Jaiswal[1], Shalini Singh[2], Srishti Jaiswal[3], Shivam[4] and Umang[5]
[1-5]ABES Engineering College, Ghaziabad, India
Email: sukjais7120@gmail.com, shalini027.singh@gmail.com, srishtijaiswal2823@gmail.com, shivamindian701@gmail.com, umangkumar20003@gmail.com

***Abstract*— Precision farming is paramount in the context of responding to world food security amidst the climatic conditions, optimization of farm input like the fertilizers to achieve the maximum harvest and minimization of damages to our environment. The standard practices of controlling nitrogen that follows rigid cycles or easy rules, are likely to lead to draining of nutrients, soil erosion and poor harvests which enhances pollution and loss of farm owners. In the framework proposed in this paper, a combination of DRA and the Gym-DSSAT crop simulation interface is suggested, and nitrogen scheduling is presented as a series of decisions in the MDP framework. Three variants of DRL are PPO-based and SAC variants along with DQN-based variants, which are trained and tested against conventional fixed-schedule baselines in various experimental settings of growing maize. SAC performed better as it recorded a 13% positive growth in yield, a 32.4% reduction in nitrogen leaching as compared to a normal practice. This means that SAC can help farmers to achieve more production of maize at the cost of the environment that is more efficient than the conventional vehicles. Such a solution of DR will introduce self-sustaining farming systems that will make farm places more resilient, profitable, and eco-friendly.**

***Index Terms*— reinforcement learning, fertilizer optimization, deep q-network, precision agriculture, decision support system for agrotechnology transfer, markov decision process, soft actor critic, proximal policy optimization.**

I. INTRODUCTION

The world population is growing at a high rate; this has necessitated a massive demand in production of sustainable food which has strained food systems. Application of such fertilizers as nitrogen (N) is important in enhancing the crops. Wu et al. [1] introduced optimization of nitrogen depending on deep reinforcement learning and suggested that improper or excessive application of such fertilizers are disastrous to the environment e.g. contamination of ground water by nitrates leaching, greenhouse gas emissions as well as soil acidification. The under-fertilization on the other hand compromises the crop health and crop yield, which affects the economic sustainability of the agricultural practices. The traditional approach to managing the fertilizers is highly dependent on general rules or inertial rule-like systems which are not dynamic and stochastic like the agricultural systems.

Grenze ID: 01.GIJET.12.2.401_24

Sutton et al. [2] addressed the fact that there exist facets such as variability in the weather on a day-to-day basis, variability of soil moisture and growth phases of crops that require an adaptive decision-making plan. The new opportunities for real-time, adaptive control of inputs into crop systems present the possibilities of machine learning and AI, and DRL in particular, outside of fixed predictions of supervised models. Although other techniques like supervised learning have already been employed in anticipating crop productivity and the spotting of diseases, they are not adequate in the real-time decision making of the agricultural operation in the changing environment of real-life agriculture. The branch of ML, which is in particular applicable to this challenge, is Reinforcement Learning (RL). A teaching experience is somewhat similar to RL: an agent can also learn optimally through experimenting with different actions in the environment, and observing the result of those actions. As time goes by, it gets to know the most prudent decisions by attempting to get the highest possible reward such as trial and error. The idea of crop management as proposed by Gautron et al. [3] presented a set of decisions whereby the RL agents are able to acquire complex strategies to conform to the real-time environment factors. This paper gives a comprehensive research on optimization of fertilizer management using deep reinforcement learning (DRL). In order to train and test DRL agents, a high-fidelity simulation environment of crop, which is called Gym-DSSAT interface, is used. We developed a formulation of fertilizer management as a Markov Decision Process (MDP) having multi-objective trade-off of level of yield and environmental cost. Three DRL algorithms DQN, PPO and SAC are run and compared, and the system proposed is tested and compared to the fixed-schedule farming practices used traditionally to demonstrate that it is more efficient and sustainable.

Part II gives a summary of the relevant literature that has gone before and the MDP is formulated in Part III. The results can be discussed and compared in the Sections IV and V respectively.

II. Literature Review

Machine learning has turned conventional systems into precision farming systems that are data driven. The most common application of Reinforcement Learning (RL) in agriculture has been curriculum control, to control irrigation and fertilizers. In contrast to the case of supervised learning where the dataset must be marked, RL is a learner-interactive model: the input to the RL agent is the environmental condition e.g., soil moisture and crop stage, the action is a management decision e.g., amount of fertilizer, and the output is a reward e.g., yield gain or penalty i.e. cost or leaching. It is based on this that Overweg et al. [4] used tabular Q-learning to train simple irrigation scheduling, despite the issue of large-dimensional state space being observed. Deep Learning along with Reinforcement Learning addresses the issue of complex and continuous state spaces, Deep Q-Networks (DQN) are chosen in cases where the decisions to be made are discrete at the fertilizer level, and the replay buffer training with an updated target network periodically, as introduced by Mnih et al. [5], can be used to overcome training instability caused by the episodic, non-smooth rollouts of crop simulators. The fractionated surrogate aim of Proximal Policy Optimization (PPO) averts damaging updates to the policy, and is best adapted to continuous dosage control, as shown by Schulman et al. [6], whereas the entropy regularization of Soft Actor-Critic (SAC) encourages more exploration of the fertilization policy space, especially in escaping local optima within early growing-season episodes [7]. Because real-world training cannot be done due to the long cycle of crop growth and the possibility of crop failure, simulation-based training using the Gym-DSSAT OpenAI Gym wrapper has become the standard benchmark to compare RL algorithms in maize and wheat production scenarios [8], and RL agents trained in this context have generally outperformed rule-based heuristics by adaptively responding to weather forecasts and changing soil conditions [9]. A head-to-head comparison conducted in identical conditions confirmed that PPO is better in the single-task nitrogen fertilization, whereas DQN shows a relative advantage in the more complex joint water-nitrogen management task, indicating that the selection of algorithms should be based on the particular structure of the management issue [10]. One consistent weakness of these methods is the fact that they assume complete state observability, as some of the variables of interest in Gym-DSSAT (including plant water stress index and daily nitrogen denitrification) cannot be directly measured in the field; modeling the problem as a Partially Observable Markov Decision Process (POMDP) and using Recurrent Neural Networks (RNNs) to represent observation histories has been found to produce stronger and Design of reward function has also evolved significantly: Baja et al. [12] proposed a restricted RL formulation that includes a specific Nitrogen Use Efficiency (NUE) reward and action constraint, showing that policies that explicitly aim at NUE avoiding the soil nutrient mining that is not well reflected by yield-based reward signals. Expanding the environmental perspective, a Probabilistic Deep Learning model within a POMDP-based DRL model concurrently imposed penalties on nitrate leaching and N_2O emissions - estimated with a 95% confidence interval coverage of 0.937- establishing that sole relying on leaching as the environmental criterion alone was not enough to manage fertilizers responsibly [13]. Most recently, a better DRA agent with an additional Bidirectional LSTM and attention-based temporal state representation was trained on a site-calibrated DSSAT environment using historical weather data (2014-2020) and the learned policy was found to be able to generalize to unseen meteorological conditions in 2022-2024 cotton field experiments, indicating that lightweight, deployable controllers on edge

Fig. 1 demonstrates the fundamental reinforcement learning loop used in our work. The DRL agent (DQN, PPO, or SAC) interacts with the Gym-DSSAT crop simulation environment at all times. Each action is generated by the agent at each timestep– a choice of the level of fertilizer. The environment contains two forms of feedback, a 12-dimensional state vector giving the present soil, weather, and crop conditions, and a reward signal that is the ratio between the crop yield obtained and the nitrogen leached. Such feedback is used to make the agent revise its policy and make more effective decisions on fertilization during subsequent timesteps. This is done thousands of times in order to enable the agent to learn a good, adaptive fertilization policy without agronomic rules being programmed into it.

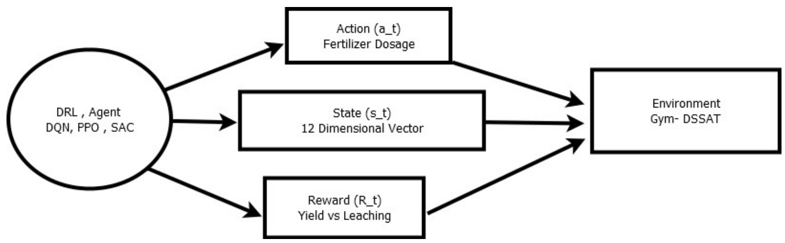

Figure 1. Reinforcement Learning Framework Diagram

Fig. 2 shows the whole procedure of the planned DRL- based fertilizer optimization structure. In parallel, three groups of state inputs are gathered, namely weather variables (solar radiation, temperature, rainfall), soil variables (water content and inorganic nitrogen), and crop variables (growth stage, biomass, and the leaf area index) and reduced and integrated into a 12-dimensional state vector. The DRL agent – SAC, PPO, or DQN – produces either continuous fertilizer dosage of 0–200 kg/ha with SAC/PPO, or a five-level discrete choice with DQN. The implementation of this action is used in the maize simulation environment of the Gym-DSSAT. The environment then calculates a multi-objective reward based on the maximization of incremental gain in biomass, minimization of total fertilizer used, and minimization of nitrate washed into the ground water. The subsequent state and policy update signal are fed back to the agent refining its decision- making through backpropagation. The agent is trained through experience to the optimal approach to fertilization in the process of repeated attempts thousands of times.

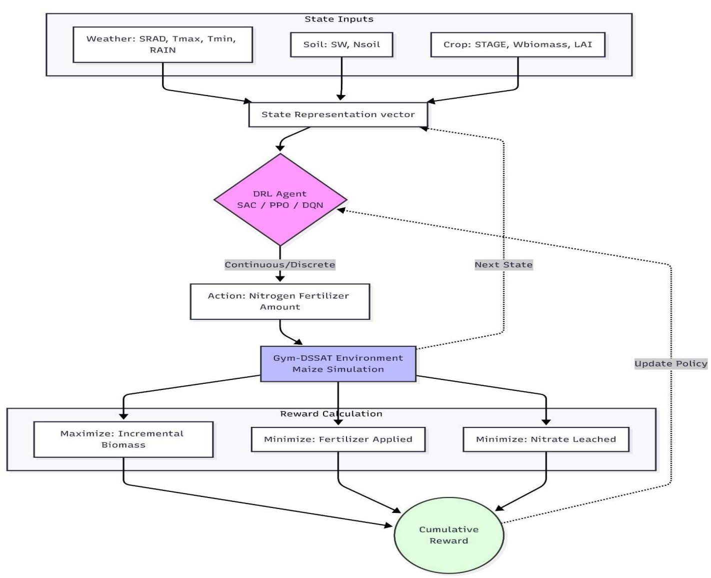

Figure 2. Detailed DRL-Based Fertilizer Optimization Process

III. PROPOSED METHODOLOGY

In the proposed framework, a Deep Reinforcement Learning technique is applied to maximize the application of nitrogen fertilizer. The system is coupled to the Gym-DSSAT environ- ment which models the maize behavioral dynamics in respect to changing weather and soil conditions. The model repre- sents the process as a finite-horizon Markov Decision Process (MDP) which is defined by the following tuple (S, A, P, R, γ), with S being the state space, A being the action space, P being the transition probability, R being the reward function and γ being the discount factor.

A. State Space

The state vector st at time step t represents the state of crop and the environment. An overall list of characteristics that are applicable to the uptake of nitrogen was chosen: Weather factors were maximum temperature (Tmax), minimum temperature (Tmin), daily solar radiation (SRAD), and rain (RAIN), Soil variables were soil water content (SW) depths and soil inorganic nitrogen (Nsoil), and Variables of crop were stage of plant growth (STAGE), plant biomass (W-biomass), and leaf area index (LAI). The dimension of the state vector is st ∈ R12, scaled so that it does not fluctuate when training the neural network.

B. Action Space

The action at represents the quantity of nitrogen fertilizer that is used on day t. In order to fit various algorithm requirements: In the case of DQN, the action space will be discretized into 5 levels: {0, 40, 80, 120, 160} kg/ha, and in the case of PPO and SAC, the action space is continuous, which is restricted to [0, 200] kg/ha, which in turn enables the dosage to be controlled accurately.

C. Reward Function

Reward function R(st,at) plays a very important role in steering an agent towards the intended goal. The multi objective rewarding function will be formulated to maximise the yield and to minimise fertilizer cost and environmental impact (nitrate leaching):

$$Rt = wl\,\Delta Yt - w2Ft - w3Lt \qquad (1)$$

ΔYt is the incremental gain in the biomass (proxy of yield) at step t; Ft is the level of fertilizer applied (kg/ha); Lt is the percentage of nitrate washed away (kg/ha); and w1,w2,w3 are weighting coefficients that trade off. To highly penalize environmental pollution, we used w1=1.0, w2=0.5 and w3=2 in our experiments.

IV. RESULT ANALYSIS

This section gives the comparative analysis of the implemented DRL algorithms with a baseline of a 'Standard Practice' approach. The routine leads to a 40-day post planting schedule, where 150 kg/ha of N is used. DQN, PPO and SAC cumulative reward curves with a 2000 training episode each show some significant information about the convergence of the algorithms like DQN, PPO and SAC. DQN was characterized by large variance of the first few episodes and slower convergence because of the discrete action space constraint, while PPO attained consistent learning but it reached a local optimum where fertilizer application was conservative. On the other hand, SAC had the quickest convergence and overall cumulative reward which was successful in balancing exploration and exploitation.

In fig. 3, with a final cumulative reward of about 92, SAC (green) has the highest final cumulative reward, and the convergence rate of the SAC (green) is significantly faster than the other algorithms, which means it has better learning efficiency. By comparison, PPO (orange) has a more stable yet slower convergence and ends up with a final reward of about 63, and DQN (blue) has the slowest convergence and ends up with a final reward of about 53, probably because of its restriction to a discrete action space. Despite the fact that all three algorithms initialize their reward values with the same value at the first episodes, they differ significantly in the way they perform over time with SAC outperforming both PPO and DQN in terms of convergence time and reward maximization. Table I offers a comparative analysis of four fertilizer management practices, SAC, PPO, DQN[14] and Standard Practice, in major performance indicators such as crop yield, fertilizer use, nitrogen leaching, and nitrogen use efficiency (NUE). Quantitatively, SAC has the highest yield being 9315 kg/ha which is 13.0 percent higher than Standard Practice (8240 kg/ha) and also the yield of SAC reduces the nitrogen leaching by 32.5 percent (42.5 kg/ha to 28.7 kg/ha). PPO is close behind with a yield of 9120 kg/ha, 10.7% above Standard Practice, and competitive NUE (72.6 kg/ha), a bit less than SAC (72.7 kg/ha). DQN offers moderate results, with a yield of 8850 kg/ha (7.4% growth) and a decrease in leaching of 10.1%. Qualitatively, the better performance of SAC can be explained by the entropy-based exploration mechanism, which allows optimizing policies better and prevents premature convergence. PPO is optimally stable because of its clipped objective function yet loses optimality to stability. Conversely, the discrete action space of DQN limits the ability to control the fertilizer application finely, as the performance becomes suboptimal. Moreover, SAC has a superior trade-off of productivity and environmental sustainability as it can maximize yield and reduce nitrogen loss at the same time.

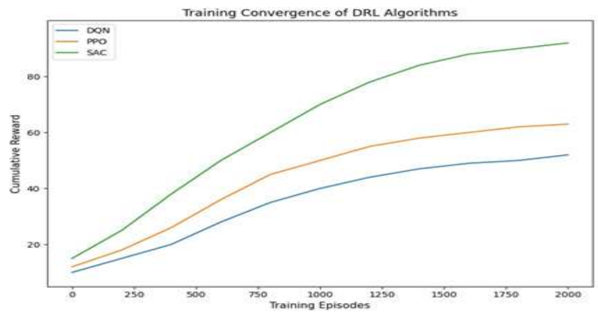

Figure 3. Training Performance over 2000 Episodes

TABLE I. PERFORMANCE COMPARISON OF ALGORITHMS

Method	Average Yield kg/ha	Fertilizer kg/ha	Leaching kg/ha	NUE kg/ha
Standard Practice	8240	150.0	42.5	54.9
DQN	8850	140.0	38.2	63.2
PPO	9120	125.5	31.0	72.6
SAC	9315	128.0	28.7	72.7

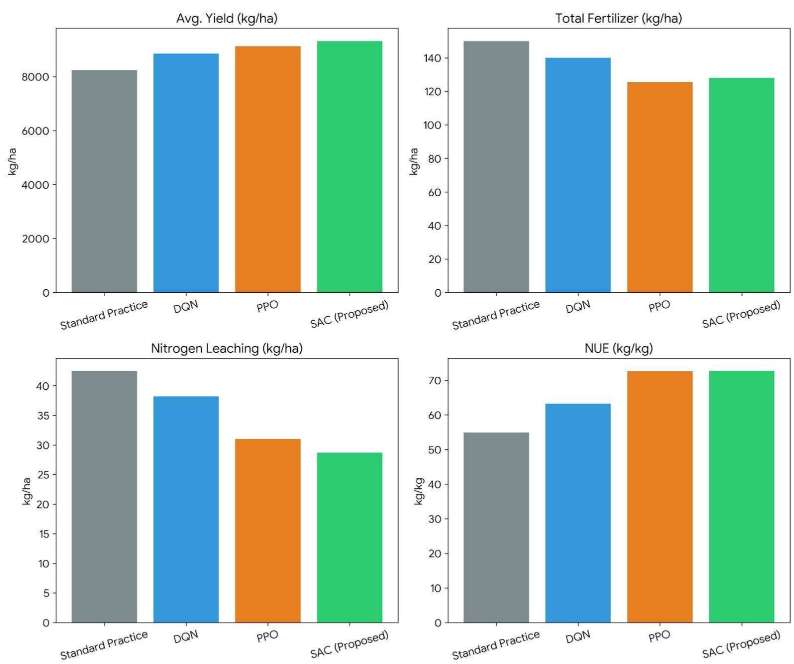

Figure 4. Comparative Evaluation of Standard Practice, DQN, PPO, and SAC across Avg. Yield, Total Fertilizer Used, N-Leaching, and NUE

Fig. 4 shows four bar graphs which are all aimed at comparing the performance of Standard Practice, DQN, PPO and SAC on four major metrics. Regarding average yield, SAC records the highest value (~9,315 kg/ha) which increases in that order Standard Practice, DQN and PPO. Compared to Standard Practice and DQN, SAC and PPO use much less total fertilizer, which is an indication of more efficient management of nitrogen. The graph of leaching of nitrogen indicates that a definite downward trend exists between Standard Practice and SAC, with the latter causing the minimal environmental pollution of 28.7 kg/ha — a decrease of 32.4 percent as compared to the baseline. Lastly, PPO and SAC are almost equal and do significantly better in terms of nitrogen use efficiency than DQN and Standard Practice. Collectively, the graphs prove that SAC provides the optimal overall balance —greater crop, less fertilizer spoilt and less leaching. Higher nutrient efficiency — it is the most effective and environmentally sustainable method of all the methods tried. The application of fertilizers was significantly different in timings of the approaches that were evaluated. When using the traditional approach to farming, the nitrogen fertilizer is normally put in a single large application at the beginning of the growing season, usually leading to the occurrence of a lot of nitrogen leaching since the crop is not in a position to take up all the amount once it has been applied. On the contrary, the policy adopted by the SAC-based agent spread fertilizer in a more strategically located crop cycle. Instead of a single application, the agent used a split application pattern, with smaller dosage of fertilizers being applied in the maize when it was at critical developmental stages, such as, the V6 stage and the VT stage when nutrient needs of the maize are high. This is quite similar to the advice of agronomic scientists, although this expertise in domain was not directly programmed into the model, which illustrates how the RL agent could be able to acquire latent biological connections between crop growth and nutrient accessibility.

V. Conclusion and Future Scope

The proposed framework outperformed the standard practice by 13% in yield and a 32.4% decrease in nitrogen leaching by modelling nitrogen fertilization as a Markov Decision Process and training agents in the Gym-DSSAT crop simulation platform by using less total fertilizer. Among the three algorithms used, SAC was the most effective, and it was found to have better exploration as measured by entropy regularization to prevent early local minima, but PPO was competitive in terms of nitrogen use efficiency and DQN was competitive approaching its discrete action space — still performed better than the baseline. That the system is capable of learning to split-apply fertilization strategy such as that provided by expert agronomic advice, without prior information of the kind ever being written in the code, reveals the true potential of reinforcement learning to simulate complicated biological interactions between crop development and nutrient dynamics. The framework has lots of space to grow in several ways in future. It is possible to add various weather records, soil profiles, and types of crops to the Gym DSSAT environment to encompass more of the variability of agriculture. The state representation can be improved by incorporating multispectral and hyperspectral images to identify the plant health and impending shortcoming. The model can also be extended to multi-crop systems, where there is a concomitant fertilization decision process of wheat, rice, and other staple crops. The state vector could include dynamic regional weather forecasts that would be used to implement fertilization strategies more likely to be timely and profit oriented components. This can be incorporated in the reward function by including market prices and input costs so that it makes the system economically workable to the smallholder farmers. Lastly, the implementation of trained policies on lightweight IoT-based edge-based actuation devices in real time, and a crowd-sourced data platform between farmers, agronomists, and researchers in different agro-climatic areas would serve a great role in enhancing the accuracy, general ization, and reliability of the system.

References

[1] J. Wu, R. Tao, P. Zhao, N. F. Martin, and N. Hovakimyan, "Optimizing nitrogen management with deep reinforcement learning and crop simulations," CVPR Workshops, pp. 1712–1720, 2022.

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

[3] R. Gautron, E. Padoux, P. Preux, and O. Maury, "gym-dssat: A crop model turned into a reinforcement learning environment," arXiv preprint arXiv:2207.03270, 2022.

[4] H. Overweg, H. N. Berghuijs, and F. A. Amani, "CropGym: a reinforcement learning environment for crop management," arXiv preprint arXiv:2104.04326, 2021.

[5] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529–533, 2015.

[6] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv:1707.06347, 2017.

[7] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft actor-critic: Off policy maximum entropy deep reinforcement learning with a stochastic actor," International Conference on Machine Learning, pp. 1861–1870, 2018.

[8] M. Turchetta, S. G. K. K. K., and A. Krause, "Learning long-term crop management strategies with CyclesGym," NeurIPS, 2022.

[9] R. Tao, P. Zhao, J. Wu, N. F. Martin, and N. Hovakimyan, “Optimizing crop management with reinforcement learning and imitation learning,” AAAI Conference on Artificial Intelligence, 2023.
[10] R. Diwakar, A. Choudhary, A. Kumar, J. Vashistha, G. Kumar, and A. K. Sahoo, “Smart Crop Recommendation System with Plant Disease Identification,” 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT), 2024.
[11] J. Balderas, D. Chen, Y. Huang, L. Wang, and R.-C. Li, "A comparative study of deep reinforcement learning for crop production management," Smart Agricultural Technology, vol. 10, p. 100863, 2025. doi: 10.1016/j.atech.2025.100863
[12] H. Yuan et al., "Learning-based agricultural management in partially observable environments subject to climate variability," arXiv preprint arXiv:2401.01273, 2024.
[13] H. Baja, M. G. J. Kallenberg, H. N. C. Berghuijs, and I. N. Athanasiadis, "Adaptive fertilizer management for optimizing nitrogen use efficiency with constrained reinforcement learning," Computers and Electronics in Agriculture, vol. 237, p. 110554, 2025. doi: 10.1016/j.compag.2025.110554
[14] S. Khaki et al., "Reinforcement learning-based agricultural fertilization and irrigation considering N2O emissions and uncertain climate variability," AgriEngineering, vol. 7, no. 8, p. 252, 2025. doi: 10.3390/agriengineering7080252
[15] Singh, Shalini, Dev Baloni, and Manash Sarkar. "Unveiling the web of privacy: A comprehensive review of threats to information security in social networks." Computational Methods in Science and Technology (2024): 185-191.
[16] B. Shang et al., "Smart irrigation scheduling for crop production using a crop model and improved deep reinforcement learning," Agriculture, vol. 15, no. 24, p. 2569, 2025. doi: 10.3390/agriculture15242569