Njiru, Ruth Dionisia Gicuku; Dong, Changxing; Appel, Franziska; Balmann, Alfons

Strategic bidding behaviour in agricultural land rental markets: Reinforcement learning in an agent-based model

International Food and Agribusiness Management Review

Provided in Cooperation with:
Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Halle (Saale)

Suggested Citation: Njiru, Ruth Dionisia Gicuku; Dong, Changxing; Appel, Franziska; Balmann, Alfons (2025) : Strategic bidding behaviour in agricultural land rental markets: Reinforcement learning in an agent-based model, International Food and Agribusiness Management Review, ISSN 1559-2448, Brill, Leiden, Vol. 28, Iss. 2, pp. 392-422, https://doi.org/10.22434/ifamr.1126

OPEN ACCESS

International Food and Agribusiness Management Review
Volume 28, Issue 2, 2025; DOI: 10.22434/IFAMR.1126

Received: 1 May 2024 / Accepted: 7 February 2025

Strategic bidding behaviour in agricultural land rental markets: reinforcement learning in an agent-based model

RESEARCH ARTICLE

Ruth Dionisia Gicuku Njiru[a]ⓘ, Changxing Dong[b], Franziska Appel[b], Alfons Balmann[c]

[a]Doctoral Researcher, [b]Doctor, [c]Professor, Department of Structural Change, Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Theodor-Lieser-Str. 2, 06120 Halle (Saale), Germany

Abstract

Agricultural land markets are crucial for efficient land allocation, yet they face complexities arising from land characteristics and the heterogeneous nature of market participants. This study explores how to address heterogeneity in the modelling process for land markets models by integrating Deep Reinforcement Learning (DRL) into the agent-based model AgriPoliS, to model strategic bidding behaviour. The simulations demonstrates that a DRL agent adapts its bidding strategies based on long-term growth objectives, experience, competitive interactions and adaptive decision-making leading to increased land rental and farm growth compared to a standard agent using a fixed bidding strategy. The results reveal how strategic behaviour not only improve individual farm performance but also affect neighbouring farms, emphasizing the dynamic interactions within land markets. By capturing the agent's strategic behaviour, this work contributes towards more realistic modelling of agricultural land market dynamics and offers insights into the implications of potential land market regulations. Future research will explore multi-agent frameworks to further refine these interactions and address the limitations of static bidding strategies.

Keywords: AgriPoliS, agent-based modelling, bidding strategy, deep reinforcement learning, farm growth, strategic interactions
JEL codes: C63, D21, Q18

ⓘCorresponding author: njiru@iamo.de

1. Introduction

The primary function of agricultural land markets is to facilitate land ownership, utilization, and access, thereby efficiently allocating the scarce resource of land and enabling investment and development for farms (De Janvry *et al*., 2001). However, land markets are highly complex due to the unique characteristics of land such as immobility, non-renewability and heterogeneous qualities (soil quality, productivity, location). Heterogeneity of the actors in terms of farm size, resources, motivation, access to information and decision-making processes further adds to the complexity (Margarian, 2014).

Because of these complexities, there is debate about how well land markets function, resulting in increased calls for land market regulations by market participants and policy makers. In Germany, for example, several federal states have formulated proposals aimed at limiting land concentration per owner or farm and controlling sale and rental prices (Deutscher Bundestag, 2018; Landtag von Sachsen-Anhalt, 2020; MLUK, 2023a; NASG, 2017). These policy initiatives, which have not yet come into force, aim to address issues such as the level of land prices, the allocation and distribution of land – including intra- and inter-sectoral distribution – and structural problems within farms, which include medium- and long-term effects on the efficiency of the sector, the distribution of rents and the exploitation of market power in land transactions. In addition to the institutional framework, the potential impact of these or similar regulations depends heavily on the structure and dynamics of competition on the land markets and on the individual behaviour and objectives of market participants. But there are still lingering questions such as whether these measures are effective or work as intended? And how can we analyse them?

These multifaceted complexities make empirical analysis of land markets challenging as addressed by Balmann *et al*. (2021), who discusses different models that deal with different aspects of the land market: Spatial competition models, for example, consider the immobility of land and concentrate on the geographical distribution of supply and demand, accounting for factors such as transportation costs and location preferences. Search and matching models factor in heterogeneity of land quality and the transaction costs of costly search and negotiation processes. Auction theory considers potential market power, which arises even without an explicit negotiation process due to the typically low number of potential bidders. However, Balmann *et al.* (2021) emphasize the challenges associated with modelling the function of land markets. Capturing all spatial, temporal, and especially behavioural aspects of land market interactions is inherently complex. One aspect that is particularly difficult to depict in such models are the various actors and the effects of their perceptions and actions. Methods that account for the impact of different actors and their perceptions and behaviour, are mostly limited to experimental insights. Buchholz *et al*. (2022) highlights that heterogeneity among farmers affects their decision making in agricultural land markets and their response to land market changes. Appel and Balmann (2023), for example, hint towards a strong influence of specific actors on the land markets. This underscores the need to better account for the heterogeneity of actors in models for analysing land markets and the assessment of policy proposals aiming at a stronger regulation of land markets.

In this respect, prior successful applications show that agent-based models (ABMs) like AgriPoliS (Agricultural Policy Simulator; Happe *et al*., 2006) can provide important insights for policy assessments. Agent-based models explicitly focus on modelling the interactions among farms (e.g. via land markets) to study emergent properties on the system level. While Heinrich *et al*. (2019) use AgriPoliS to specifically analyse various types of land market regulations. Other applications of AgriPoliS such as Happe *et al*. (2008), Uthes *et al*. (2011) and Appel *et al*. (2016) implicitly address land market implications of policy reforms, as they assess the effects of different policy measures or changes on agricultural structures, which are fundamentally linked to land market interactions.

Agent-based models are flexible regarding modelling of agent behaviour. Examples of behavioural approaches range from simple rules to computational intelligence, including learning. However, agent-based models of the agricultural sector often assume that farm behaviour is driven by profit- or utility-maximizing principles,

portraying farmers as perfectly rational price-takers (AgriPoliS, cf. Happe *et al.* (2006); MP-MAS, cf. Berger and Schreinemachers (2006); Schreinemachers and Berger (2011); SWISSLand, cf. Möhring *et al*. (2016)). These approaches often also overlook the complexities of decision-making in real-world scenarios. Common weaknesses include the sensitivity of optimization results to uncertain expectations, neglect of strategic considerations, and the assumption of perfect rationality among agents.

With this paper, we focus on how incorporating the behaviour of strategically oriented agents – who make their bidding decisions based on long-term growth objectives, past experience, current farm conditions, competitive interactions, and adaptive decision making – can improve the modelling of land market dynamics. To that end, we explore how the integration of Deep Reinforcement Learning (DRL) into the decision-making of the agents in the existing agent-based model, AgriPoliS, enables strategic decision making in land markets. Such an approach is original for the analysis of agricultural land markets and related policies. To the best of our knowledge, there is no agent-based model in the agricultural sector using DRL (cf (Groeneveld *et al*. (2017); Huber *et al*. (2018); Kremmydas *et al*. (2018); Storm *et al*. (2020)).

The paper is structured as follows: Section 2 focuses on selected concepts related to ABM and DRL. Section 3 illustrates AgriPoliS and the experimental set-up for integrating DRL in AgriPoliS. In Section 4, the results of the simulation are presented and thereafter discussed in Section 5. In Section 6, a conclusion of the paper is presented.

2. State of the art

2.1 Agent-based models and their behavioural approaches

ABM is a bottom-up approach for simulating complex and dynamic systems through modelling the behaviour and interactions of entities referred to as agents (Crooks and Heppenstall, 2012). The agents could represent individual or collective agents in pursuit of a specific objective(s). The agents are autonomous, heterogeneous, active and interact with each other and their environment. Through the individual behaviour and interactions, emergent phenomena and system dynamics are observed (Bonabeau, 2002; Railsback and Grimm, 2019).

ABMs usually employ behavioural approaches that are prevalent in agricultural policy analyses, such as myopic optimization using mixed-integer programming. Kremmydas *et al.* (2018) discovered in a literature review that approximately 45% of modelling frameworks explicitly employ mathematical programming optimization, including alternative methods like positive mathematical programming. Moreover, approximately 30% of models rely on simple rules, while 25% are based on behavioural heuristics. Similar findings by Groeneveld *et al.* (2017) for agent-based land-use models indicate widespread use of optimization, heuristics, and stochastic decision-making components. Further, An (2012) offers a specific methodological classification of behavioural models for ABMs, examining coupled human and natural systems. This classification includes microeconomic models, space theory-based models, psychosocial and cognitive models, institution-based models, experience- or preference-based decision models, participatory ABM, empirical or heuristic rules, as well as evolutionary programming and assumptions, and calibration-based models. However, these behavioural approaches, especially those used in models for the agricultural sector, do not consider the individual strategic behaviour of farms.

An explorative study by Appel and Balmann (2023) emphasizes the effects of individual behaviour on land markets. They analyse the spatial influences of different behavioural clusters of farm managers. As a further aspect, Appel and Balmann (2023) conclude that the development and actions of a farm are not only influenced by other actors on the local land market, but also by the irreversibility of decisions and interactions. Their findings also align with the findings from Shang *et al.* (2021) where technology adoption and diffusion is influenced by the farmer's behaviour, market conditions, institutional frameworks and social networks. Capturing these spatial, behavioural and temporal aspects of land market interactions is inherently complex.

Machine learning methods (ML) have shown a lot of promises in modelling complex behaviour and have the potential to be used to capture land market interactions.

2.2 Machine learning and artificial intelligence used in agriculture

Storm *et al.* (2020) offer an overview of machine learning (ML) approaches in agricultural and applied economics, emphasizing quantitative analysis. They also discuss the use of ML for surrogate models, which approximate the mapping between inputs and outputs of complex underlying models, such as learning to replicate behavioural characteristics of agent-based models (Shang *et al.*, 2024). Similarly, van der Hoog (2017) explores the utilization of artificial neural networks as surrogate models or meta-modelling approaches in ABMs to reduce complexity and computational demands. Additionally, both Storm *et al.* (2020) and van der Hoog (2019) mention the potential application of ML and reinforcement learning (RL) in simulation models, enabling agents to learn optimal behaviour in dynamic, reactive environments.

In recent years, deep learning approaches based on complex multi-layer artificial neural networks, known as deep neural networks (DNNs), have been combined with RL. Such DRL approaches have proven exceptionally successful in mastering complex strategic games like Go and Chess. Prominent examples include AlphaGo, AlphaGoZero and AlphaZero (Silver *et al.*, 2018; Silver *et al.*, 2017).

Studies show that DRL can improve the precision of behavioural modelling in ABMs by allowing the agents to modify their behaviour through interactive learning and interaction with their environment thus generating more flexible and adaptive agents who interact with their environment in such a way that results in optimal behaviour and thus optimization agents (Dehkordi *et al.*, 2023; Osoba *et al.*, 2020; Turgut and Bozdag, 2023; Zhang *et al.*, 2023). For instance, Vargas-Pérez *et al.* (2023) demonstrated that DRL agents outperformed static agents as an aid in building a decision support system for the best media advertising investment strategy. Olmez *et al.* (2022) show that DRL proves useful in creating agents that display intelligent and adaptive behaviour through time and space in a predator and prey ABM. Li *et al.* (2019) applies an extended Roth-Erev RL algorithm (Roth and Erev, 1995), for individual agent decision-making process in a residential land growth ABM which significantly improved the agents' adaptive behaviour and improved the models' simulation power. Liang *et al.* (2020) adopted a DRL algorithm for bidding strategies in electricity generating companies while accounting for incomplete information and in high dimensional continuous state and action spaces.

RL enables agents to learn which actions (such as bids on the land market) lead to greater long-term rewards (such as increased equity capital) through repeated interaction within the environment. Such a RL framework is described by Sutton and Barto (2018): It is based on states, actions, transitions and rewards. In a simplified RL setup, the algorithm is expressed as a Markov decision process (MDP), represented by a tuple (S, A, T, R, π, γ), as elaborated below:

- States (S), where $S = s_1, s_2, s_3, \ldots, s_n$, is a finite set of all possible situations that the agents may find themselves in within their environment.
- Action $a_1, a_2, a_3, \ldots, a_n$, is a set of all possible actions that the agents may take within their environment based in their current state (s) at every possible time step (t).
- Transition (T) is the transition function between states. Therefore, when an agent takes an action (a_t) at a given timestep (t), it transitions from the old state (s_t) to a new state (s_{t+1}) in the environment.
- Reward (R) is the reward resultant in the agent's action (a_t) and transitioning to a new state (s_{t+1}). It is usually represented as a scalar reward and can be either negative or positive. The reward can be given at the end of every time step or can be given after the end of several steps. The goal of the agent is to maximize the cumulative reward which is the summation of all the rewards received until the terminal time step.
- RL policy function π defines how the agent chooses the action to take in its current state to maximize its cumulative reward i.e. maps that states into action. Discount factor γ is a factor between 0 and 1

that discounts future rewards that helps the agent balance between short-term and long-term rewards. Values close to 0 indicate actions that lead to long-term rewards. And for a policy π, we can estimate a value function V^{π}, which is the value of the accumulated reward.

In a simplified way, as depicted in Figure 1, the agent observes the current state of its environment and uses this information to decide which action to take. The agent then receives new information and the reward because of the action. Based on the new observation, the agent decides whether to take new action or repeat the action. The cycle continues until the terminal state. The goal of the agent is to learn the optimal RL policy from its environment. One critical aspect of RL is that the agents learn from exploring its environment but at the same time exploiting good actions that have been taken before i.e. trade-off between exploration and exploitation. In the next section we delve a bit deeper on the model AgriPoliS and the proposed framework for the integration of DRL in AgriPoliS.

3. Methodology and experimental set-up

3.1 AgriPoliS

AgriPoliS (Agricultural Policy Simulator) is an ABM used to simulate effect of diverse policies and regulations on agricultural structural change over time (Balmann, 1997; Happe *et al.*, 2006). The AgriPoliS environment is a virtual landscape with spatially located agricultural farms which are represented as farm agents. The farm agents are closely similar to typical farms in the region, heterogenous, have different factor endowments, different managerial skills, different farm ages, pursue a defined goal e.g. income maximization and exhibit myopic behaviour (Appel *et al.*, 2016; Balmann, 1997; Happe *et al.*, 2006; Sahrbacher *et al.*, 2012). The farms are defined prior to initialized based on real farm data, European Union's Farm Accountancy Data Network (FADN), handbook data on farming practices (e.g., for Germany, Association for Technology and Construction in Agriculture (KTBL)), farm structural survey (FSS) and/or expert knowledge (Njiru *et al.*, 2024; Sahrbacher and Happe, 2008).

The land market is at the centre of AgriPoliS where rental plots become available upon expiration/termination of existing rental contracts, farms downsizing or farm closures. Farms can solely grow through renting additional land through the land rental auction market. The rental market also forms the basis for interaction among the agents through their competition for additional land. The rental market takes place at the beginning of the production period. The rental plots are spatially distributed and the farm agent incurs transport costs between their own farm plots and the plots available (Happe, 2004; Kellermann *et al.*, 2008). The farms present bids to the land rental market. The agent with the highest bid receives the plot. The auction is held in an iterative manner until all the plots are allocated. The bids (equation 1) reflect the shadow price, *q* (additional benefit

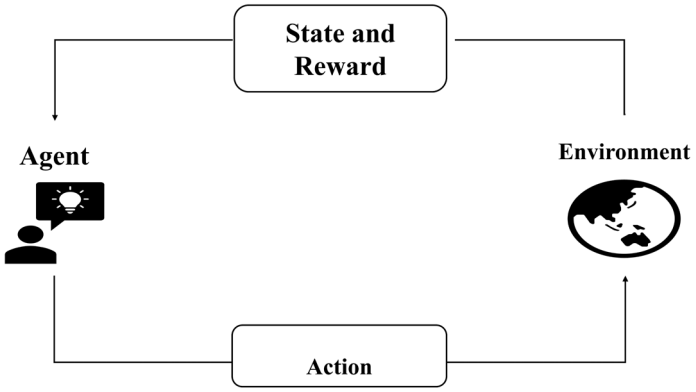


Figure 1. Simplified reinforcement learning framework. Adapted from Sutton and Barto (2018).

of renting the land), the spatial location of the plot (calculated by the transport cost between the farm plots and the plots available for rent), TC, and a fixed land rental coefficient, β, which determines the share of the shadow price that is passed on to the landowner (Happe, 2004). The remaining share 1-β remains with the farmer to cover further costs (including taxes, administrative expenses, fees) and the risk. For β the following applies: $0< \beta <1$. In a perfectly competitive market, one would expect β to be 1.0. However, a theoretical study by Graubner (2018) emphasizes the potential for spatial price discrimination in the agricultural land market. Accordingly, in regions where space plays a significant role, such as those with larger farms, one would assume the share transferred to rental prices tends to be lower. In the present study, AgriPoliS simulates a region with a limited number of rather large farms. Therefore, β is assumed to be 0.5 for all farms of the respective model region.

$$\text{bid} = [q - \text{TC}] * \beta \quad (1)$$

Through use of mixed integer programming (MIP), factor endowments (land, labour (family labour, hired labour), fixed assets), key production activities (livestock, crops), investment options (machinery, livestock housing), financing options (short-term and long-term credit, liquidity) and other activities specific to the region (e.g. manure disposal, livestock density restrictions) are represented simultaneously while considering resource constraints inherent in agriculture. In each period, the farm agents follow a sequential process i.e. participates in the land rental auction market, makes investment, decides on what to produce, does the actual production, does the annual farm accountancy, decides on whether to continue or exit farming and the process restarts. A typical simulation run in AgriPoliS is done over 25 iterations/time periods.

As the agents are myopic, they only consider decisions one period ahead. This poses significant risks for the farm agents such as jeopardising the farms' financial stability, increased vulnerability to economic shocks and reduced resilience to challenges such as changing market conditions. This necessitated the extension of the AgriPoliS model towards strategic behaviour through strategic bidding decisions that factor in long-term planning while leveraging on past experiences (e.g. previous rental rates), current farm situation (e.g. Liquidity), consideration of competitive interactions (e.g. number of farms) and adaptive decision-making.

In our novel approach, the farm agent would come up with a strategic bidding DRL policy based on state variables reflecting farm and sector conditions, bid interdependence, and long-term planning. The reward mechanism is based on cumulative sum of equity capital at the end of all the iterations. In the next sections, details on the methodological and experimentation setup of integrating DRL in AgriPoliS are explained.

3.2 The framework for integrating DRL in AgriPoliS

In this new framework, one agent is equipped with DRL while the other agents used the AgriPoliS behaviour. The DRL agent formulates a bid i.e. action based on the current state of the environment which is transmitted to the land market and subsequently receives feedback i.e. new states and uses this information to formulate a new bid. The iterative process continues until the end of the simulation run, at which point the cumulative sum of equity capital is calculated. The technical details are discussed in the following sections while access to the ODD+D protocol, datasets and code are available on the AgriPoliS website.

3.2.1 The objective

In AgriPoliS, every farm agent makes decisions independently and interacts with other agents through different markets, of which the land market is the most important. As the decisions are made by optimizing the profit only for the current year, they are myopic and might not be optimal for the agent's long-term development. The objective is to find a bidding strategy in the land market that maximizes a single farm agent's long-term equity capital by enhancing the agent with reinforcement learning ability and therefore making strategic bidding decisions in the land rental markets that maximizes the agents' long-term growth.

3.2.2 The environment

As this work serves mainly as proof of concept, the investigated region is deliberately made small. It consists of seven typical farms from the agricultural region, Altmark in Germany. One of the farms is designated as the DRL agent while the other 6 farms are standard AgriPoliS farms. The framework consists of three logical components as shown in Figure 2. The first component is the machine learning unit (MLU), through which the bidding strategy can be learnt. The second component is the adapted AgriPoliS environment where the simulation take place, henceforth referred to as the APS-ENV. The third component is a message queue system that allows reliable communication between the MLU and the APS-ENV both locally and remotely and is referred to as the COM-MQ.

This modular framework design allows the independent development of the MLU with respect to not only the different training algorithms and parameters but also the programming languages and computer operating systems.

An agent in AgriPoliS henceforth referred to as the DRL agent is trained with the MLU to formulate strategic bids also known as actions which are then transmitted into the APS-ENV via COM-MQ. The resultant data i.e. states and rewards are transmitted back to the MLU through the COM-MQ. Essentially, in this framework, the actions go from MLU through COM-MQ to APS-ENV while the states and rewards flow in the opposite direction.

3.2.3 The state space

The state space reflects the current state of the farm and region, the interdependence of bids on other farm level decisions (e.g. how higher/lower bids affect the investments) and long-term planning effect. The selected variables to represent a DRL agent's state are shown in Table 1. There are two different soil types (arable and grazing) and 47 different investment options. A state has 67 variables because some variables are differentiated between the soil types. The agent receives state s_t and uses the information to prepare the bid.

3.2.4 The action

In the standard AgriPoliS Model, the β is fixed at 0.5 representing 50% of the valuation while for this new framework the DRL agent's β is determined by a neural-network-based DRL architecture. The action space is thus a continuous variable.

3.2.5 The reward

Since the goal is to learn a bidding strategy that maximizes the agent's long-term reward, the DRL agent only gets a reward at the terminal state. A state is terminal, if it is the state at the end of an AgriPoliS simulation

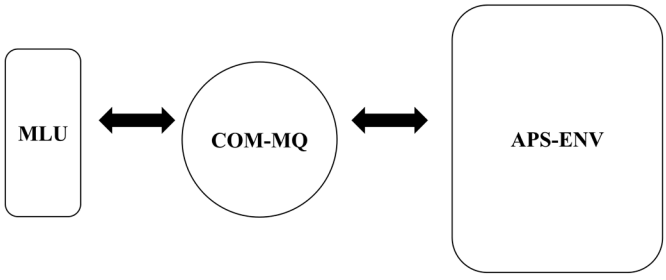


Figure 2. Reinforcement learning framework within AgriPoliS.

Table 1. Variables representing DRL agent states.

Name	Type	Number of variables	Level	Notes
Terminating plots	List of integers	10	Farm	Distribution of plot numbers over rest contract length (1 to 5 years). This is differentiated between arable land and grassland
Liquidity	Real	1	Farm	Ability to of the farm to meet short term liabilities
Farm age	Integer	1	Farm	Age of the farm
Investments	List of real	47	Farm	Remaining life of the investments
Previous rental rate	Real	2	Farm	The latest amount of rent paid by the agents. This is differentiated between arable land and grassland
Management coefficient	Real	1	Farm	This is reflecting the farms' managerial ability by manipulating the variable costs.
Free plots	Integer	2	Region	Number of available (remaining) free plots in the region. This is differentiated between arable land and grassland
Number of farms	Integer	1	Region	Number of competing farms
Average rent price	Real	2	Region	The average rental rate in the region for the previous year. This is differentiated between arable land and grassland.

run. For the experiments, the simulation run is over 10 years with the first year denoted as iteration 0 being the initialization run. The reward in this case is the cumulative sum of its equity capital over the simulation run. We use cumulative sum of equity rather than final equity to encourage consistent improvement in farm performance across iterations rather than focusing only on the final state. Emphasizing the final equity value, on the other hand, could lead to riskier strategies that prioritize highest profits in certain iteration, while potentially risk the farms stability. In addition, models trained exclusively on final equity values are expected to show increased sensitivity to the number of iterations considered. In contrast, the use of cumulative rewards mitigates this sensitivity by encouraging the agent to consider both short-term and long-term profitability, and thereby assumingly promoting a balanced and robust learning curve that includes both immediate gains and sustainable performance.

Let's denote the transition from the state s to s' by taking the action a as (s, a, s'), then the cumulative reward R can be described as

$$R(s, a, s') = \sum r_i$$

if s′ is a terminal state, otherwise $R(s, a, s') = 0$. Here r_i is the equity capital of the DRL agent after the i^{th} year in the simulation with AgriPoliS.

3.2.6 The algorithm and experimental set up

In Figure 3, the detailed processes in the three components in the framework are illustrated. All data flows through COM-MQ are indicated with blue lines, whilst the black lines show the data flows within a subsystem APS-ENV or MLU. "partial rewards" within COM-MQ means the equity capital after every single year.

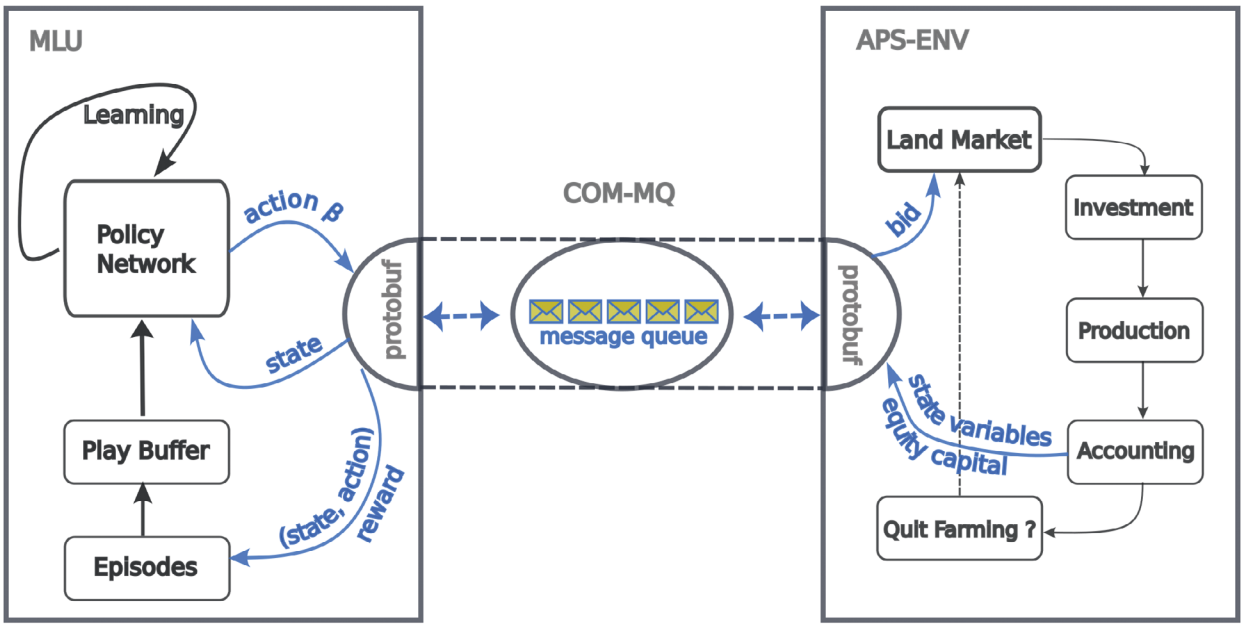


Figure 3. Detailed dataflow of DRL in AgriPoliS.

The process begins by initializing the APS-ENV, the DRL policy network with random weights, the play buffer for storing the training data, episode (the number of iterations for every AgriPoliS simulation run), episodes per epoch, the number of epochs among other parameters. A summary of values of the parameters is presented in Table 2. As the action space is continuous, the learning is approached through a deterministic policy gradient algorithm (Silver *et al.*, 2014). In this algorithm only one DRL policy network is used, the input of which is the state and output the action. The learning is iterative with fixed number of epochs. In each epoch, it goes through three steps: training data collection, network update and DRL policy testing.

To collect training data, the DRL policy network is used to obtain actions enhanced with noise to encourage exploration. Since the DRL agent only get nonzero reward at terminal states, the simulations with AgriPoliS will not be stopped or interrupted and they are rollout episodes with the same length, the number of simulation years. The reward is given at the end of a simulation/episode, i.e. cumulative sum of the equity capital. Only episodes with largest rewards are used for training. The state action pair (s, a) in these episodes are used for updating the DRL policy network. These pairs with the rewards of their episodes are put in a memory called play buffer. The play buffer is updated, if new episodes have larger rewards.

The update of the DRL policy network exploits supervised learning algorithms where states are the inputs and actions from the play buffer are target values. Concretely we use mean squared error (MSE) as the loss function and ADAM (Adaptive Moment Estimation) as the optimizer.

After updating the DRL policy network, the last step in an epoch is to test the DRL policy learned. This is a standard simulation run with AgriPoliS, obtaining the actions from the updated DRL policy network, but without additional noises. The learn behaviour can be seen in the rewards of the testing runs of all the epochs.

The algorithm combines Monte Carlo tree search (MCTS) with supervised learning like the algorithm in AlphaGo (Silver *et al.*, 2016). Although here the search tree is not explicit but implicitly saved in the play buffer.

Table 2. Training parameters for the reinforcement learning

Parameter	Value	Notes
N	2000	Number of epochs
S	30	Number of simulations (episodes) per epoch
Y	10	Number of years per simulation (episode)
*N*_hidden	2	Number of hidden layers of DRL policy network
*S*_hidden	16	Size of hidden layers of policy network
LR	1e–4	Learning rate of DRL policy network
Batch	8	Batch size

To make the algorithm more understandable, the pseudocode in Python style is given below.

Algorithm: MLU reinforcement learning algorithm

```
def initialization (N, S, Y, ENV):          # ENV is an instance of APS-ENV
        P.init()                            # initialization of DRL policy network P with arbitrary weights
        PlayBuffer=[]                       # the play buffer for training data
        noise.reset()                       # reset noise generator
        num_epochs=N                        # number of epochs
        num_simulations=S                   # number of simulations per epochs
        num_years=Y                         # number of years for every simulation with AgriPoliS
        best_reward=0                       # best episode reward
        MQ .init(ENV)                       # MQ is an instance of COM-MQ

def simulation(test=False):                 # Simulation with AgriPoliS
        y=0                                 # current year
        episode=[]
        R=0
        while y < num_years:
                state=MQ.get_state()
                action=P(state,w)           # action from DRL policy network
                if not test:
                        action += noise()
                MQ.send_action(action)
                equity=MQ.get_equity()
                R=R + equity
                if not test:
                        episode.append((state,action))
                y += 1
        return R
```

```
def main(N, S, Y, ENV):
    initialization(N, S, Y, ENV)
    e=0                                  # current epoch
    while e < num_epochs:
        s=0                              # current simulation
        while s < num_simulations:
            R=simulation()
            if R > best_reward:
                PlayBuffer.update(episode)
            best_reward=R
            s += 1
        P.update()                       # Learning with MSE loss function and ADAM
                                         # optimizer using training data from PlayBuffer
        R=simulation(test=True)   # test the learned DRL policy
        e += 1
        output (e, R)                    # output the test run
```

3.3 Evaluation

Unlike in traditional ML methods, evaluation is not based on the loss function as the optimal reward is unknown and the training data is dynamically updated, the value of the loss function varies with the epochs and therefore it is not suitable to indicate the quality of the learnt DRL policy. The DRL learnt behaviour is evaluated by the output of the cumulative sum of equity capital from the testing run in every epoch. Comparison is also made between the DRL agent and the Baseline agent in terms of rented land, bidding strategy and the cumulative sum of equity capital.

3.4 The agents set-up

There are 7 farm agents spatially located in the region (Figure 4). A description of the farm agents at initialization is provided in Table A1 in the Appendix. The farm agents are heterogeneous and stable implying relatively low risk of the farms exiting farming within a simulation since the reward is only calculated at the end of the simulation run. The farms are also active in the land market, albeit with varying levels of activity as shown in the Baseline scenario which is a standard AgriPoliS run without DRL (Figure 5).

In the first part, Farm 4 (F4) is designated as the DRL agent because it is neither too aggressive nor too passive regarding the rental market while the other farms are designated as the standard AgriPoliS agents. In the second part, the DRL experiments are repeated while setting the other farms as the DRL agents to test the flexibility and adaptability of the framework to other farms with different structures. The results from the experiments are presented in Section 4.

4. Simulation results

In this section, the results from the APS-ENV and MLU integration are presented. In section 4.1, a single DRL agent (F4) was trained to compete against 6 standard AgriPoliS agents (Baseline agents) in the land market. The behaviour of the DRL agent was evaluated based on cumulative rewards, total rented plots and the bidding strategy and then compared to the Baseline agent representing the standard AgriPoliS behaviour without DRL. In Section 4.2, the flexibility of the framework to be adapted to other agents is explored. Here, the simulations were replicated by designating different farm agents as the DRL agent and subsequently comparing the results to those of the Baseline agent.

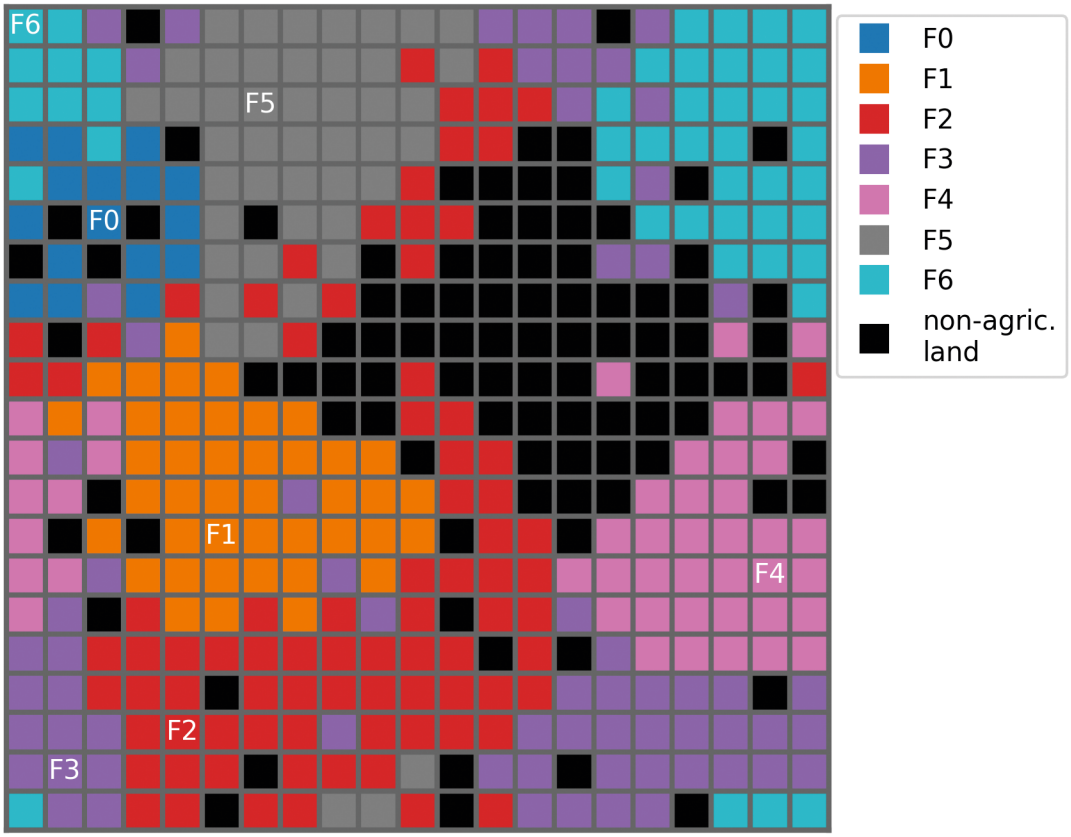


Figure 4. Regional representation of farms in AgriPoliS. Note: regions in AgriPoliS are modelled as tori, which means that neighbouring regions wrap around each other, creating a continuous loop without boundaries.

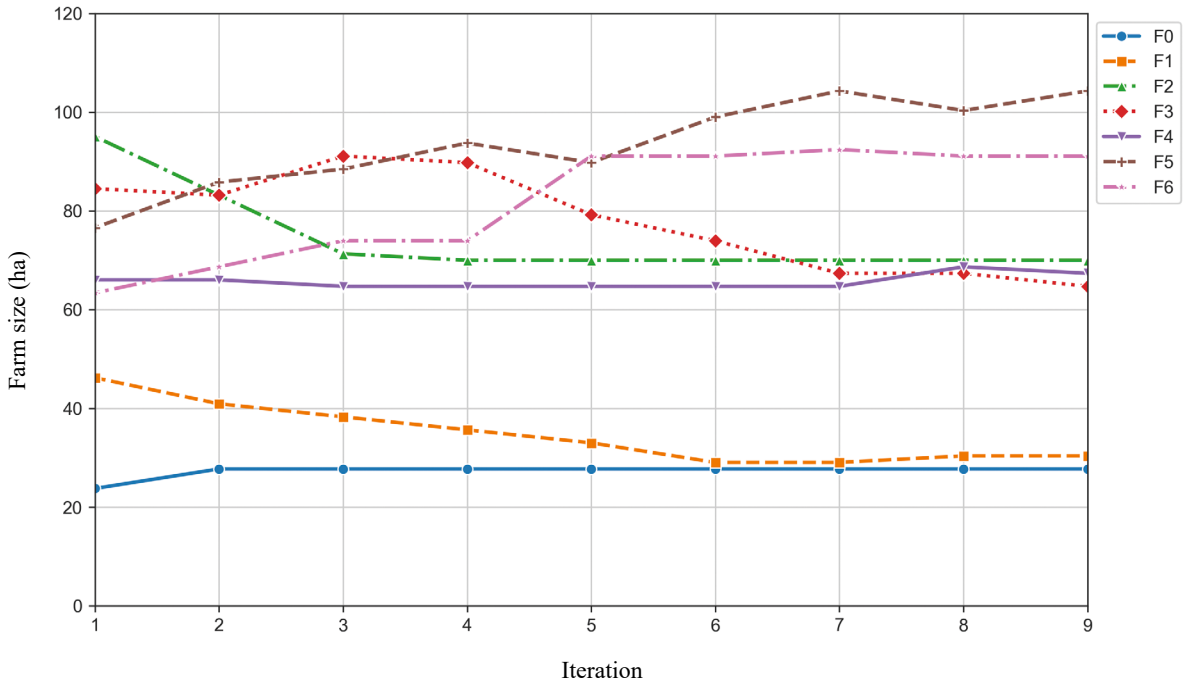


Figure 5. Baseline scenario for farm size in AgriPoliS.

4.1 Training the DRL agent

The DRL agent (F4) was trained and the learnt behaviour over 2000 epochs was generated. The DRL agent's highest cumulative sum of equity capital reflects a 7.29% increase relative to the Baseline agent's cumulative sum of equity capital (Table 3).

Figure 6 displays the cumulative sum of equity capital for the DRL agent over 2000 epochs. The graph demonstrates that the DRL agent surpassed the performance of the Baseline agent from the first epoch and continues to improve over subsequent epochs to stabilize to approximately 7.9 million euros around the 1000th epoch.

Figure 7a illustrates that the strategy resulting in the highest cumulative reward was not a fixed bidding strategy but rather that the DRL agent varied the beta coefficient between the iterations during a simulation run. Furthermore, the DRL agent's farm size increased because of renting more plots of land from the land rental market compared to the Baseline agent's farm size (Figure 7b).

The land rental market is key for interaction among the farms because one farm can only grow if other farms downsize or even exit. Consequently, it was vital to investigate how DRL affected other farms through their interaction in the land market (Figure 8).

Table 3. Comparison of the agent's equity capital (Baseline vs DRL).

	Baseline agent (F4)	DRL agent (F4)	Relative change (%)
Initial equity capital (€)	789 808	789 808	–
Annual average equity capital (€)	820 786	880 586	7.29
Cumulative sum of equity capital (€)	7 387 073	7 925 292	7.29

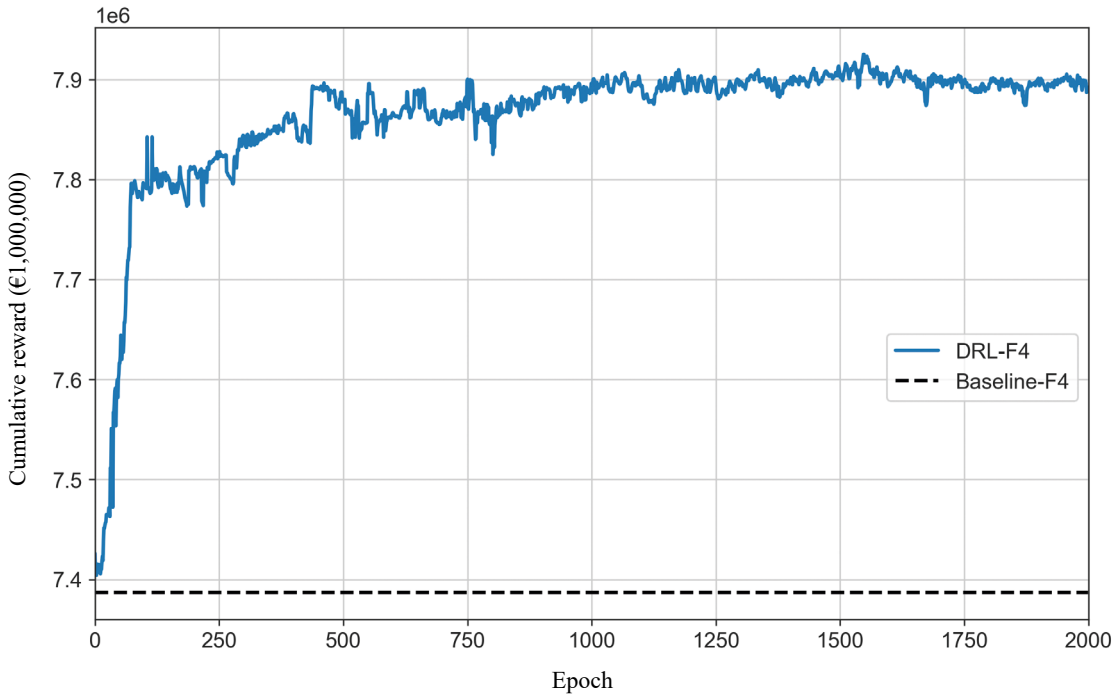


Figure 6. Cumulative reward for the DRL agent over 2000 epochs.

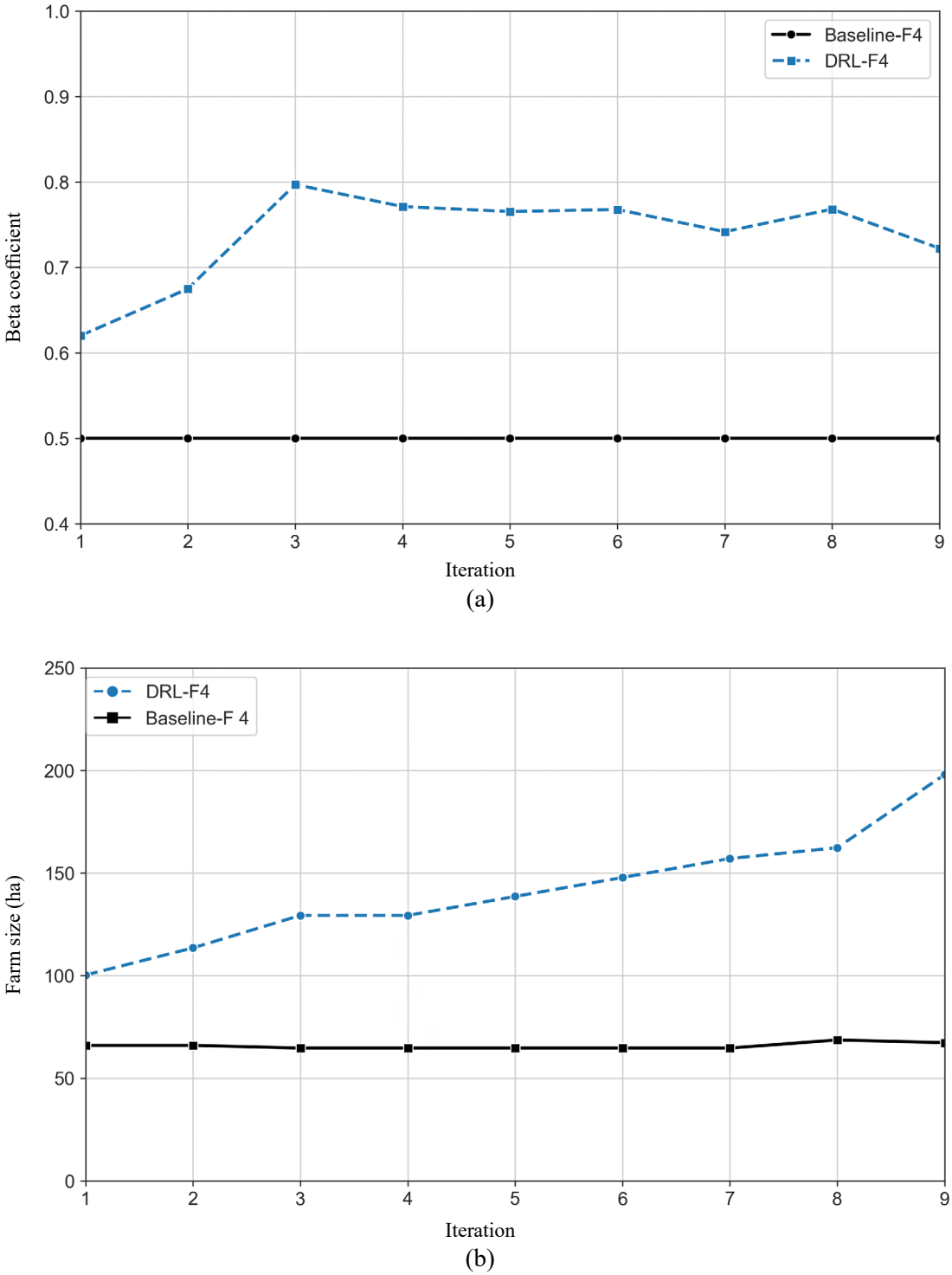
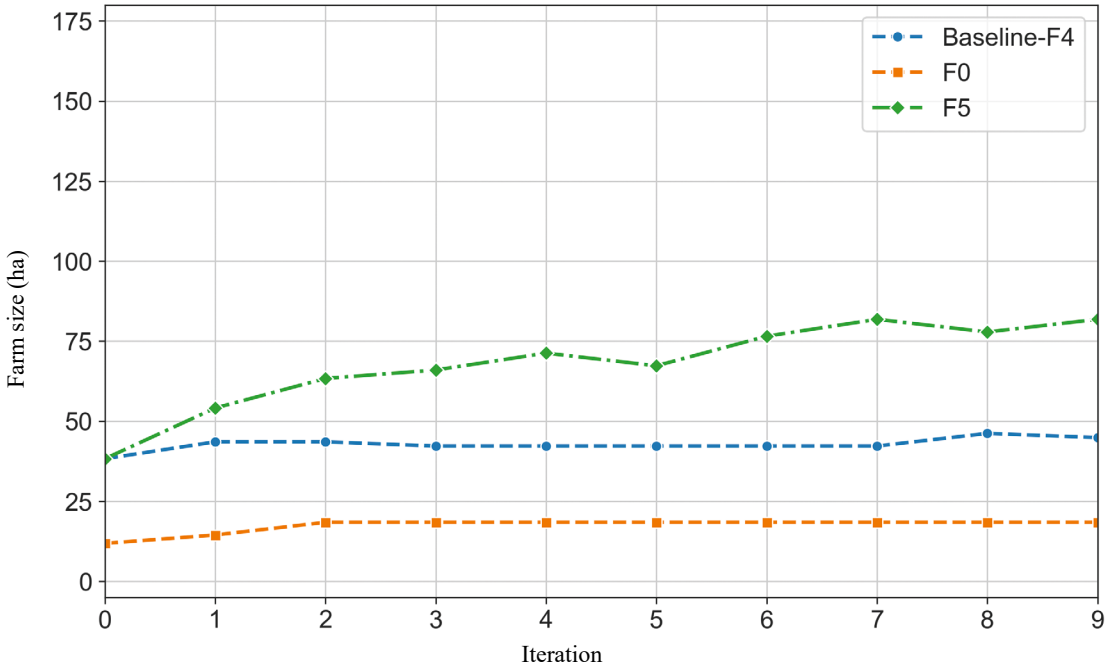
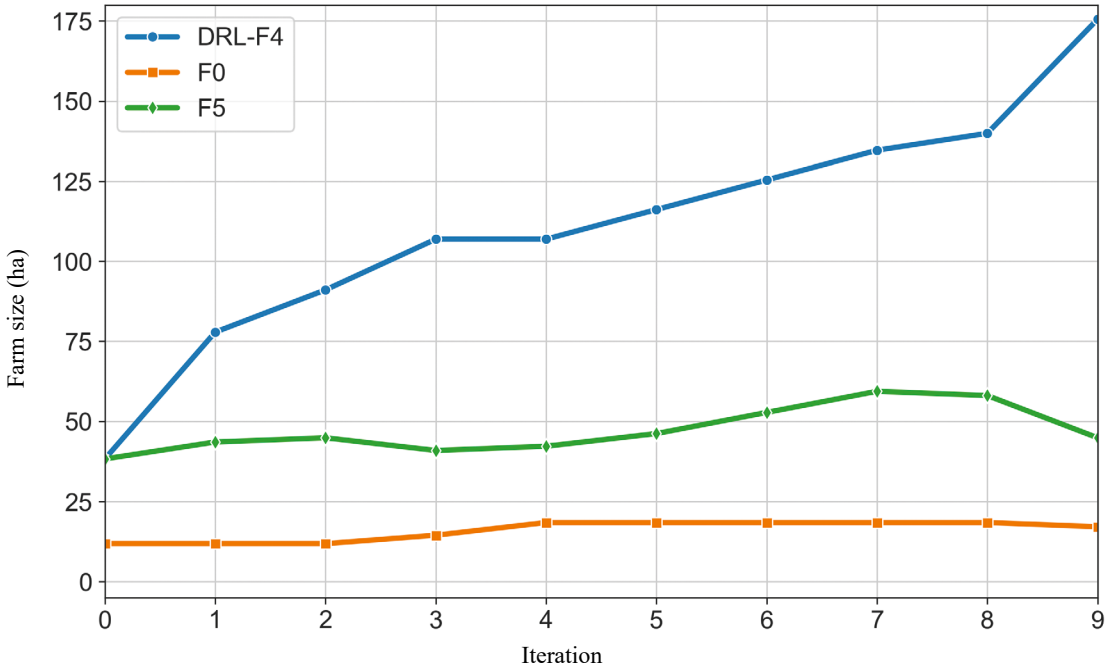


Figure 7. (a) Best bidding strategy for the DRL agent. (b) Evolution of farm size (ha).



(a) Baseline



(b) DRL

Figure 8. Effect of learning on farm size for other farms.

For example, F5 experienced a decline in the farm size in the DRL scenario compared to the Baseline scenario because of less rented plots due to the DRL bidding strategy. The size of F0 remained relatively low indicating that F0 was not that active on the land market in both scenarios. The same effect was also observed for all the farms as illustrated in Fig. A1 in the Appendix.

4.2 Flexibility of the framework

Based on the success of the training process in Section 4.1, and to test the adaptability and flexibility of the framework, experiments were repeated using the same network architecture and hyperparameters for all the other farms. The results were evaluated based on the same reward structure i.e. the farm agent's cumulative sum of equity capital and compared with the Baseline.

Based on the metrics in Table 4, it is evident that there was an increase in the cumulative sum of equity capital for all the farms in the DRL scenario compared to their Baseline except for Farm 0 which showed a slight decrease of 0.05% in the cumulative sum of equity capital with DRL. F0, F1, F2 experienced a modest increase of 2.30, 1.84 and 1.57%, respectively, in the cumulative sum of equity capital. F4 to F6 saw significant increase of 7.29, 9.18 and 9.17%, respectively, in the cumulative sum of equity capital with DRL.

Figure 9 displays the cumulative sum of equity capital at every epoch relative to the Baseline. The results across 2000 epoch demonstrate variations in the speed of learning among the farm agents. F0 however, performed worse even after learning over 2000 epochs. An investigation as to whether restricting training to 2000 epochs caused the lack of convergence was conducted. A 2-fold increase in the number of epochs did not improve the learning for the agent. This shows that the agent did not benefit from learning and this is further reinforced by the fact that the farm seemed to rent relatively the same (very low) number of rental plots in both DRL and Baseline scenarios (Figure 10a). In contrast, F6 rented significantly more land when using DRL as compared to the Baseline (Figure 10b). This also led to a significant increase in their farm size and subsequent increase in their cumulative sum of equity capital (Table 4). Also, all the other farms rented more plots of land by using DRL as compared to the Baseline. However, farms F1 to F3 showed negligible increment in rented plots between the DRL and Baseline scenarios (Figure A2 in the Appendix). This is also reflected by a relatively low increase in their cumulative sum of equity capital. Like F6, F5 rented significantly more land when using DRL as compared to the Baseline which is also in line with their significant increase of cumulative sum of equity capital.

In Figure 11, an illustration of bidding strategies for selected individual farms indicates that by varying their bidding coefficient the farm agents can maximize their rewards (see Figure A4 in the Appendix for the best bidding strategy of the farms not presented here). Every farm had a unique best bidding strategy while all the other farms maintained a fixed bidding strategy in the DRL scenario.

Table 4. Comparison of the cumulative sum of equity capital for all the farms.

Farm agent	Baseline cumulative reward	DRL maximum cumulative reward	Relative change (%)
F0	2 288 694	2 287 647	−0.05
F1	3 932 183	4 022 778	2.30
F2	6 850 762	6 976d853	1.84
F3	12 669 940	12 868 645	1.57
F4	7 387 073	7 925d292	7.29
F5	7 889 960	8 613 919	9.18
F6	8 354 720	9 120 867	9.17

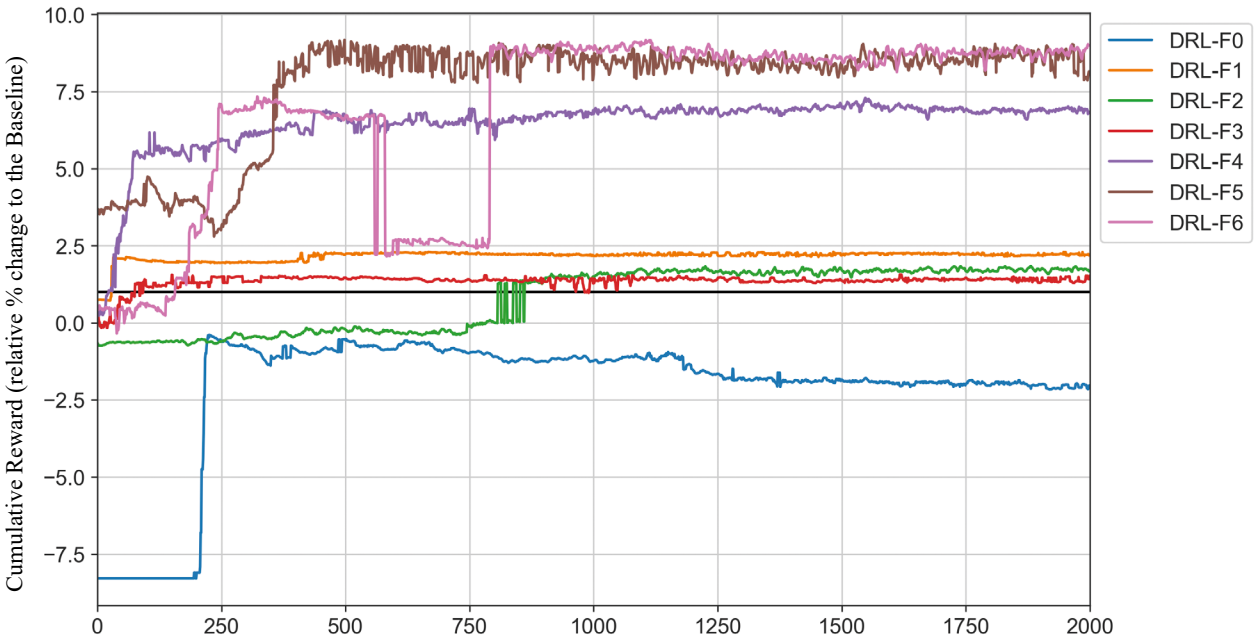


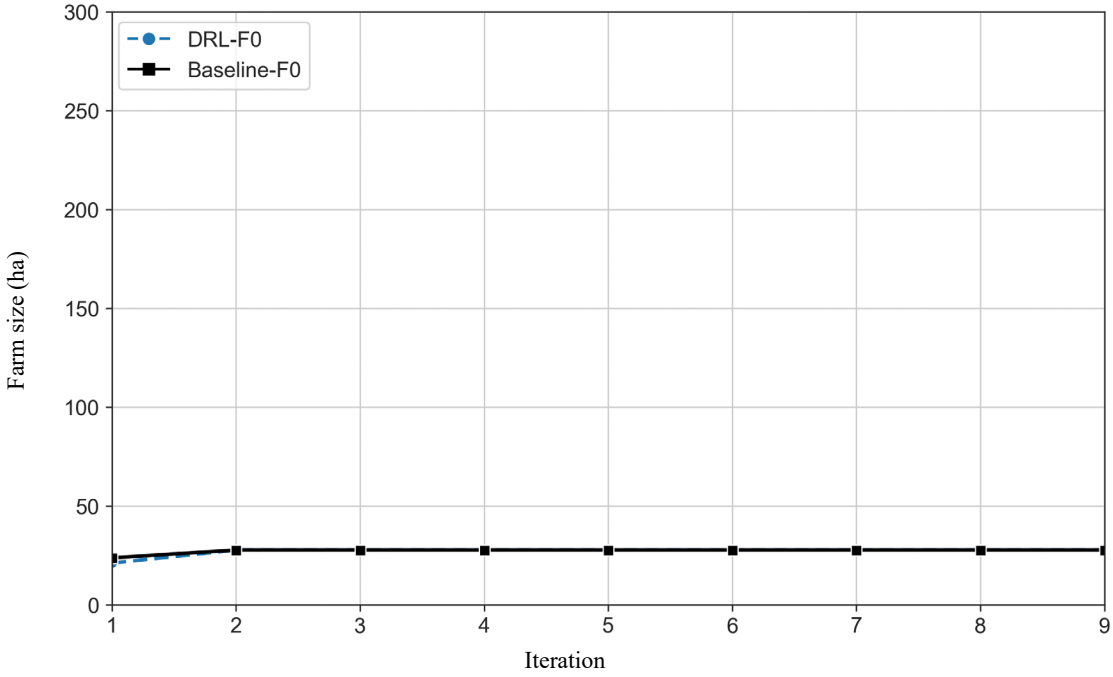
Figure 9. Relative change in cumulative reward in DRL compared to the baseline for all farms.

5. Discussion

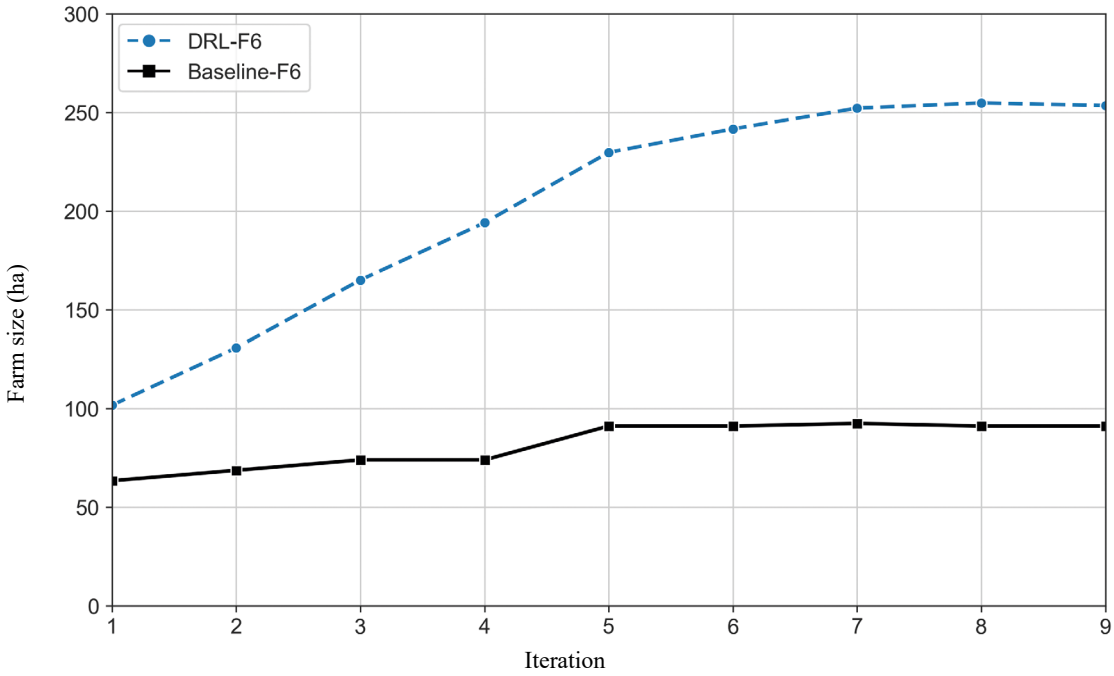
In this paper, we explored how the integration of DRL into the decision making of agents in AgriPoliS enables strategic behaviour and interactions. The exploration began by constructing an advanced framework consisting of a learning unit (MLU), communication unit (COM-MQ) and an adapted AgriPoliS environment (APS-ENV). Through the framework, a single agent learnt a bidding strategy that maximized their cumulative reward i.e. cumulative sum of equity capital based on state variables reflecting farm and regional conditions, bid interdependence, and long-term planning. The results were then compared to a Baseline scenario (standard AgriPoliS agent) where the agent used a fixed bidding coefficient. The DRL agent demonstrated superior adaptability and strategic decision-making compared to the standard AgriPoliS agent.

The strategic superiority of the farm agent is demonstrated by their greater competitiveness on the land market and stronger farm growth, as demonstrated by the increased farm size (Figure 7b). Comparatively, the DRL agent performed better than the Baseline agent as shown by the increase in the cumulative sum of equity capital (Table 3 and Figure 6). Additionally, the results indicate that the best strategy based on the state variables significantly differ from the use of a fixed bidding coefficient (Figure 7a). On the other side, the bidding strategy proved to be detrimental to other farm agents (with standard fixed β) as the DRL agent outbid them in their quest to rent more land from the land rental market (Figure 8). Overall, the DRL framework underlines how adaptive decision-making can promote a farm's long-term growth.

To further evaluate the adaptability of the DRL framework, simulations were repeated across all farm agents using the same hyperparameters, allowing us to assess its flexibility for heterogeneous farm types. The results showed that the framework could be applied effectively to other farms exhibiting different structures and financial capabilities (Figure 9). With the DRL bidding decisions being adaptive and driven by past experiences, current farm situation, regional conditions, and competitive interactions, it is therefore sensible that the bidding strategies differed across the farms further illustrating how the heterogeneity of farms leads them to different bidding strategies and thus difference in farm growth. However, at this stage, the learning processed is more effective for farm agents with high activity on the land market and not for farms with minimal land market activity (Figure 10). Although it should be theoretically possible that farms with low

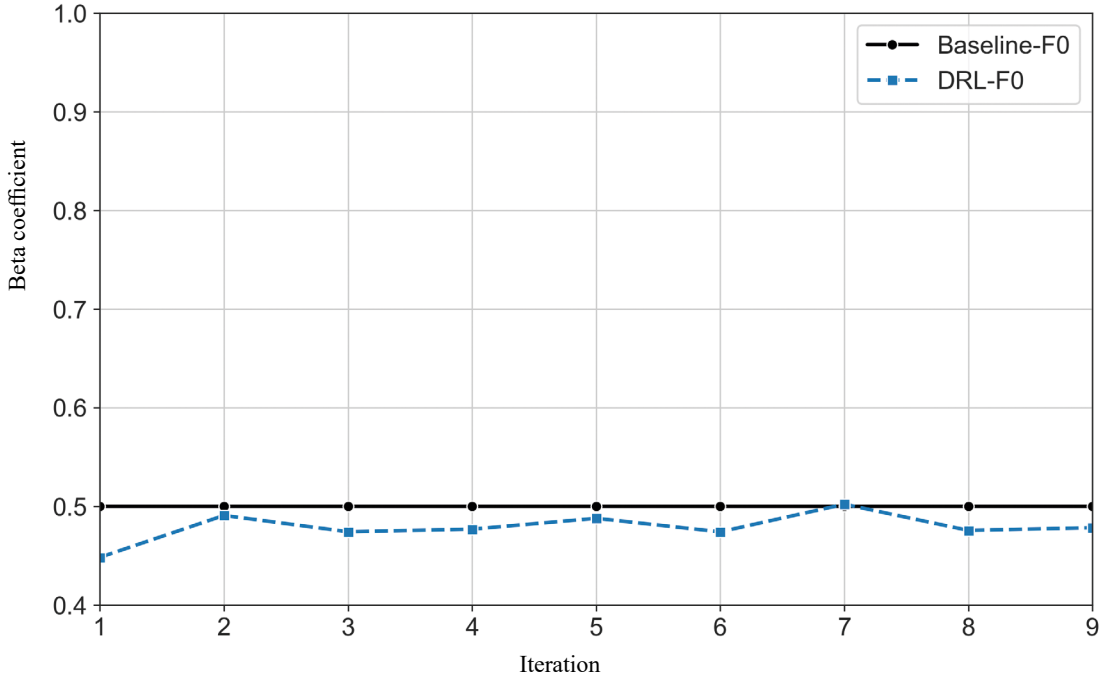


(a) Farm 0

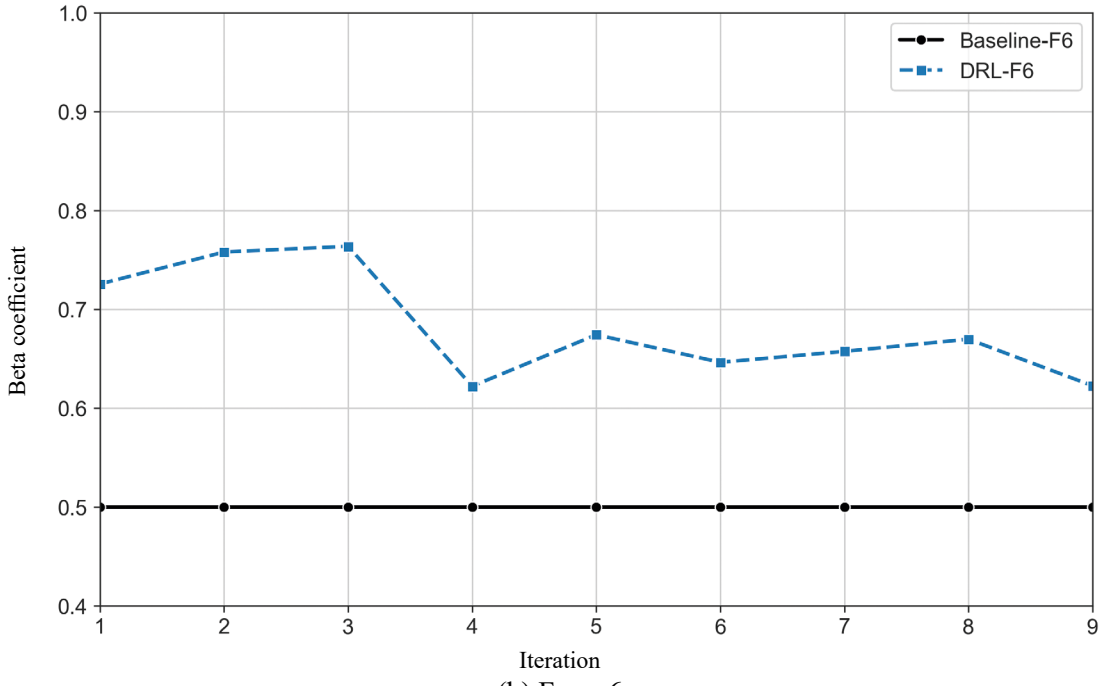


(b) Farm 6

Figure 10. Evolution of farm size(ha) for selected farms.



(a) Farm 0



(b) Farm 6

Figure 11. Best bidding strategy for selected farms.

activity on the land market could benefit from additional tuning of the hyperparameters for enhanced learning, this might not be too relevant for studying land market interactions. Land market dynamics are driven by a few very strong actors (Appel and Balmann, 2023) and therefore it should be sufficient to model improved strategic behaviour for some agents while others show the default behaviour. These results underline the potential of the DRL framework to model the dynamics of land markets more realistically by reflecting the strategic heterogeneity among farms.

Building on these findings, we can consider the broader implications of adapting standard AgriPoliS agents' bidding strategies beyond a fixed coefficient. The results could be understood as an indication that assuming a constant bidding coefficient of 0.5 for standard AgriPoliS agents might just be too low. However, Graubner (2018) emphasizes the potential for spatial price discrimination in the agricultural land markets: Due to distance costs, farms face a convex (price-elastic) regional land supply, resulting in imperfect competition and a reduced share of income from land being transferred to rental prices. This theoretical work is also underpinned by empirical studies. For instance, Kilian *et al.* (2012) examined the share of expected income from land (direct payments) capitalized into land prices, finding incidences ranging from 28 to 79% for Germany (Bavaria), depending on the direct payment type and land category. Although direct payments are only a part of the expected additional income used to determine shadow prices and, consequently, land market bids, these findings suggest that farmers typically bid only a fraction of their shadow price. Assuming an optimized fixed bidding coefficient for each standard AgriPoliS agent, we estimate these coefficients to range between 0.44 and 0.77 (see Table A2 in the Appendix). The bidding strategy employed by the DRL agent surpasses this optimized fixed coefficient by adaptively adjusting its renting coefficient to the specific individual and land market conditions in each iteration. These results indicate that strategic advantage in the land market is not simply a matter of having a higher or lower bidding coefficient, but rather lies in adaptively adjusting bidding decisions based on past experiences, current farm situation, regional conditions, and competitive interactions, resulting in optimized spatial and temporal price discrimination.

While our findings underscore the value of the DRL framework for enabling adaptive bidding strategies for a more realistic modelling of land market dynamics, certain limitations must be acknowledged in terms of the study's scope and computational feasibility. In this study, the region was modelled with 7 farm agents, whereas a typical real-world region, would comprise several hundred farms, creating a more dynamic environment and requiring simulations that could extend over several days, weeks or even months. This simplification helped to reduce complexities such as training time, computational cost (memory, processing power) and convergence to a stable solution. Future studies could aim to model larger regions by addressing these complexities through approaches like parallelized training, automated hyper-parameter optimization and use of faster GPUs.

A key criticism of DRL is that the agent learns through repeated simulated interactions with the environment, allowing them to experience millions of possible scenarios – an advantage not available to real-life farmers who face external constraints and limited information. However, in this framework, the information provided to the DRL agent (Table 1) mirrors what a well-informed and well-connected regional farm manager might realistically possess. Just as a farm manager would rely on a combination of experience, knowledge, and intuition to refine their bidding strategies over time, the DRL agent refines its bidding strategies over time based on experience and evolving expertise. These aspects of experience and intuition cannot be captured by traditional normative models (Buchholz *et al*., 2022). While DRL cannot explicitly cover all these aspects either, it enables agents to learn optimal strategies through iterative adjustments, creating a form of experience or intuition. Furthermore, our analysis (Figure 6) illustrates that the DRL agent effectively leveraged the available information to make competitive decisions, resulting in increased cumulative equity capital from the outset. This suggests that strategic behaviour can emerge even without extensive information (Silver *et al.*, 2021). While indeed DRL has the theoretical potential to provide virtually unlimited information through extensive iterations and data, our study was constrained by computational limitations, prompting us to only provide information in Table 1 to the DRL agent which does not completely reflect the complexity of human decision making.

Another critical aspect of the presented DRL framework is the assumption that other agents adopt a fixed bidding strategy throughout the simulation. This simplification creates a relatively stationary environment, allowing the DRL to exploit the static behaviour of the other agents. In reality, however, other farms would adjust their behaviour in response to unsuccessful bids. These adjustments would have implications at both the farm and regional levels, introducing an additional layer of modelling complexity that has not been effectively addressed yet. To capture this additional layer of complexity, future work will focus on advancing the current DRL framework towards a multi-agent deep reinforcement learning (MADRL) approach. In such an advanced framework, multiple agents would use DRL to determine their bidding decision in the land market, rather than relying on a rule-based heuristic like the fixed bidding coefficient. MADRL is particularly useful for modelling more complex interactions, competition, and dependencies among agents. With MADRL, agents can sense and respond to other agents' strategies, which may lead to higher efficiency, fiercer competition or even collaboration among them (Busoniu *et al*., 2008; Osoba *et al*., 2020). This approach would enable us to study optimal land allocation among farm agents, potential improvements of land use and investment planning, and possible increases in earnings. Through improved behavioural strategies, it could also contribute to discussions on how farms may adapt to, or even circumvent land market regulations. Despite these advantages, MADRL introduces certain challenges, particularly in managing the complexity of a continuously changing environment: as all agents learn simultaneously, the environment becomes non-stationary making it more challenging for the agents to strategically adjust their actions. Additionally, as the number of agents increases, the 'curse of dimensionality' intensifies, with a corresponding rise in the number of state and action variables to consider – introducing computational challenges that we are actively working on.

6. Conclusion

The aim of this paper was to explore how farms strategic decision-making behaviour in land markets could be captured more accurately by integrating DRL as an alternative behavioural approach in AgriPoliS. In this framework, a single DRL agent learns adaptively from the environment, enabling them to generate bids, focusing on long-term growth. The agent's decisions are informed by state variables that capture the farm's status, regional conditions, and competitive interactions thus simulating aspects of long-term planning, experience, and adaptive decision making in the land markets. Within this set-up, the agent competed against the standard myopic behaviour of traditional AgriPoliS agents.

The experiments demonstrated that the DRL agent was more competitive in the land market and managed to increase their long-term growth as indicated through increases in farm size and cumulative sum of equity capital. As the DRL agent's farm expanded through strategic bidding, other farms faced a corresponding decrease in rented land. Additionally, when different farms designated as the DRL agent, they adopted unique bidding strategies resulting in increased, though varied, growth for the respective farm. These results underscore the potential of using DRL in capturing strategic decision-making for heterogeneous farms, laying the foundation for further exploration of strategic farm behaviour in agricultural modelling. The findings provide valuable insight into improving potentials in modelling of farm decisions and behaviours, addressing a gap where most models rely on rule-based heuristics or myopic decision making in land markets.

Future work will expand the model to include a larger region with more farms, introducing greater competition and improving the realism of the representation of regional farm structures. Additionally, we plan to extend the framework to MADRL, enabling multiple agents to use DRL to learn and adjust their strategies over time. Such a model could serve as a valuable tool for modellers, practitioners and policymakers, helping to explore land market dynamics and assess the effects of proposed land market regulations.

Acknowledgements

This work was financially supported by the German Research Foundation (DFG) through Research Unit 2569 "Agricultural Land Markets – Efficiency and Regulation".

References

An, L. 2012. Modeling human decisions in coupled human and natural systems: Review of agent-based models. *Ecological Modelling* 229: 25–36. https://doi.org/10.1016/j.ecolmodel.2011.07.010

Appel, F. and A. Balmann. 2023. Predator or prey? Effects of farm growth on neighbouring farms. *Journal of Agricultural Economics* 74(1): 214–236. https://doi.org/10.1111/1477-9552.12503

Appel, F., A. Ostermeyer-Wiethaup and A. Balmann. 2016. Effects of the German Renewable Energy Act on structural change in agriculture – the case of biogas. *Utilities Policy* 41: 172–182. https://doi.org/10.1016/j.jup.2016.02.013

Balmann, A. 1997. Farm-based modelling of regional structural change: A cellular automata approach. *European Review of Agricultural Economics* 24(1): 85–108. https://doi.org/10.1093/erae/24.1.85

Balmann, A., M. Graubner, D. Müller, S. Hüttel, S. Seifert, M. Odening, J. Plogmann and M. Ritter. 2021. Market power in agricultural land markets: concepts and empirical challenges. *German Journal of Agricultural Economics (GJAE)* 70(4): 213–235. https://doi.org/10.30430/gjae.2021.0117

Berger, T. and P. Schreinemachers. 2006. Creating agents and landscapes for multiagent systems from random samples. *Ecology and Society* 11(2).

Bonabeau, E. 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences* 99(suppl 3): 7280–7287. https://doi.org/10.1073/pnas.082080899

Buchholz, M., M. Danne and O. Musshoff. 2022. An experimental analysis of German farmers' decisions to buy or rent farmland. *Land Use Policy* 120: 106218. https://doi.org/10.1016/j.landusepol.2022.106218

Busoniu, L., R. Babuska and B.D. Schutter. 2008. A Comprehensive Survey of Multiagent Reinforcement Learning. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)* 38(2): 156–172. https://doi.org/10.1109/TSMCC.2007.913919

Crooks, A.T. and A.J. Heppenstall. 2012. Introduction to Agent-Based Modelling. In A.J. Heppenstall, A.T. Crooks, L.M. See and M. Batty (eds). *Agent-Based Models of Geographical Systems*. Springer, Dordrecht, pp. 85–105. https://doi.org/10.1007/978-90-481-8927-4_5

Dehkordi, M.A.E., J. Lechner, A. Ghorbani, I. Nikolic, E. Chappin and P. Herder. 2023. Using machine learning for agent specifications in agent-based models and simulations: A critical review and guidelines. *Journal of Artificial Societies and Social Simulation* 26(1). https://doi.org/10.18564/jasss.5016

De Janvry, A., E. Sadoulet and W. Wolford. 2001. *Access to land and land policy reforms*. Vol. 3. UNU World Institute for Development Economics Research, Helsinki.

Deutscher Bundestag. 2018. *Antrag – Bodenmarkt transparent gestalten und regulieren – Eine breite Eigentumsstreuung erhalten – Bäuerlichen Betrieben eine Zukunft geben*. Drucksache 19/5887. https://dserver.bundestag.de/btd/19/058/1905887.pdf

Graubner, M. 2018. Lost in space? The effect of direct payments on land rental prices. *European Review of Agricultural Economics* 45(2): 143–171. https://doi.org/10.1093/erae/jbx027

Groeneveld, J., B. Müller, C.M. Buchmann, G. Dressler, C. Guo, N. Hase, F. Hoffmann, F. John, C. Klassert, T. Lauf, V. Liebelt, H. Nolzen, N. Pannicke, J. Schulze, H. Weise and N. Schwarz. 2017. Theoretical foundations of human decision-making in agent-based land use models – A review. *Environmental Modelling and Software* 87: 39–48. https://doi.org/10.1016/j.envsoft.2016.10.008

Happe, K. 2004. *Agricultural policies and farm structures – Agent-based modelling and application to EU–policy reform*. Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Halle (Saale), Germany. Available online at http://ageconsearch.umn.edu/record/14945/files/st040030.pdf

Happe, K., A. Balmann, K. Kellermann and C. Sahrbacher. 2008. Does structure matter? The impact of switching the agricultural policy regime on farm structures. *Journal of Economic Behavior and Organization* 67(2): 431–444. https://doi.org/10.1016/j.jebo.2006.10.009

Happe, K., K. Kellermann and A. Balmann. 2006. Agent-based analysis of agricultural policies: an illustration of the agricultural policy simulator AgriPoliS, its adaptation and behavior. *Ecology and Society* 11(1).

Heinrich, F., F. Appel and A. Balmann. 2019. *Can land market regulations fulfill their promises?*. FORLand–Working Paper 12. Humboldt University, Berlin, Germany. https://doi.org/10.18452/20890

Huber, R., M. Bakker, A. Balmann, T. Berger, M. Bithell, C. Brown, A. Grêt–Regamey, H. Xiong, Q.B. Le and G. Mack. 2018. Representation of decision-making in European agricultural agent-based models. *Agricultural Systems* 167: 143–160. https://doi.org/10.1016/j.agsy.2018.09.007

Kellermann, K., C. Sahrbacher and A. Balmann. 2008. *Land Markets in Agent-based Models of Structural Change*. 107th EAAE Seminar. Sevilla, Spain. Available online at https://ageconsearch.umn.edu/record/6647/files/cp08ke18.pdf

Kilian, S., J. Antón, K. Salhofer and N. Röder. 2012. Impacts of 2003 CAP reform on land rental prices and capitalization. *Land Use Policy* 29(4): 789–797. https://doi.org/10.1016/j.landusepol.2011.12.004

Kremmydas, D., I.N. Athanasiadis and S. Rozakis. 2018. A review of agent based modeling for agricultural policy evaluation. *Agricultural Systems* 164: 95–106. https://doi.org/10.1016/j.agsy.2018.03.010

Landtag von Sachsen-Anhalt. 2020. *Entwurf eines Agrarstrukturgesetzes Sachsen-Anhalt – ASG LSA*. Drucksache 7/6804. Avaialble online at https://padoka.landtag.sachsen-anhalt.de/files/drs/wp7/drs/d6804rge.pdf

Li, F., Z. Xie, K.C. Clarke, M. Li, H. Chen, J. Liang and Z. Chen. 2019. An agent-based procedure with an embedded agent learning model for residential land growth simulation: The case study of Nanjing, China. *Cities* 88: 155–165. https://doi.org/10.1016/j.cities.2018.10.008

Liang, Y., C. Guo, Z. Ding and H. Hua. 2020. Agent-based modeling in electricity market using deep deterministic policy gradient algorithm. *IEEE Transactions on Power Systems* 35(6): 4180–4192. https://doi.org/10.1109/TPWRS.2020.2999536

Margarian, A. 2014. The reflexive relationship between local land markets and farmers' strategies in Germany. *Studies in Agricultural Economics* 116(1): 1–12. http://dx.doi.org/10.7896/j.1328

MLUK (2023a). *Gesetzentwurf der Landesregierung – Gesetz zum Erhalt und zur Verbesserung der brandenburgischen Agrarstruktur auf dem Gebiet des landwirtschaftlichen Bodenmarkts*. Available online at https://mluk.brandenburg.de/sixcms/media.php/9/bbg-asg-entwurf-20230417-mluk.pdf

Möhring, A., G. Mack, A. Zimmermann, A. Ferjani, A. Schmidt and S. Mann. 2016. Agent-based modeling on a national scale–Experiences from SWISSland. *Agroscope Science* 30: 1–56.

NASG (2017): *Entwurf – Gesetz zur Sicherung der bäuerlichen Agrarstruktur in Niedersachsen*. Niedersächsisches Agrarstruktursicherungsgesetz (NASG). Available online at https://www.niedersachsen.de/download/118210

Njiru, R., F. Appel, C. Dong and A. Balmann. 2024. *Application of deep learning to emulate an agent-based model*. FORLand–Technical Paper 3. Humboldt University, Berlin, Germany. http://dx.doi.org/10.22004/ag.econ.340874

Olmez, S., D. Birks and A. Heppenstall. 2022. Learning complex spatial behaviours in ABM: an experimental observational study. *arXiv Preprint*: arXiv:2201.01099. https://doi.org/10.48550/arXiv.2201.01099

Osoba, O.A., R. Vardavas, J. Grana, R. Zutshi and A. Jaycocks. 2020. *Modeling agent behaviors for policy analysis via reinforcement learning*. In: 19th IEEE International Conference on Machine Learning and Applications (ICMLA). Miami, FL, USA, pp. 213–219. https://doi.org/10.1109/ICMLA51294.2020.00043

Railsback, S.F. and V. Grimm. 2019. *Agent-based and individual-based modeling: a practical introduction*. Princeton University Press, Princeton, NJ.

Roth, A.E. and I. Erev. 1995. Learning in extensive–form games: experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior* 8(1): 164–212. https://doi.org/10.1016/S0899-8256(05)80020-X

Sahrbacher, C. and K. Happe. 2008. *A methodology to adapt AgriPoliS to a region*. Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Halle (Saale), Germany. Available online at https://www.agripolis.org/bilder/A_Methodology_to_Adapt_AgriPoliS_2008.pdf

Sahrbacher, C., A. Sahrbacher, K. Kellermann, K. Happe, A. Balmann, M. Brady, H. Schnicke, A. Ostermeyer, F. Schönau and C. Dong. 2012. *AgriPoliS: An ODD–Protocol*. Leibniz Institute of Agricultural Development in Transition Economies (IAMO), Halle (Saale), Germany. Available online at https://www.agripolis.org/bilder/agripolis-odd-2012.pdf

Schreinemachers, P. and T. Berger. 2011. An agent-based simulation model of human–environment interactions in agricultural systems. *Environmental Modelling and Software* 26(7): 845–859. https://doi.org/10.1016/j.envsoft.2011.02.004

Shang, L., J. Wang, D. Schäfer, T. Heckelei, J. Gall, F. Appel and H. Storm. 2024. Surrogate modelling of a detailed farm-level model using deep learning. *Journal of Agricultural Economics* 75(1): 235–260. https://doi.org/10.1111/1477-9552.12543

Shang, L., T. Heckelei, M.K. Gerullis, J. Börner and S. Rasch. 2021. Adoption and diffusion of digital farming technologies – integrating farm-level evidence and system interaction. *Agricultural Systems* 190: 103074. https://doi.org/https://doi.org/10.1016/j.agsy.2021.103074

Silver, D., A. Huang, C.J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel and D. Hassabis. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529(7587): 484–489. https://doi.org/10.1038/nature16961

Silver, D., T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan and D. Hassabis. 2018. A general reinforcement learning algorithm that masters chess, shogi and Go through self-play. *Science* 362(6419): 1140–1144. https://doi.org/10.1126/science.aar6404

Silver, D., G. Lever, N. Heess, T. Degris, D. Wierstra and M. Riedmiller. 2014. Deterministic policy gradient algorithms. *Proceedings of the 31st International Conference on Machine Learning* 32(1): 387–395. https://proceedings.mlr.press/v32/silver14.html

Silver, D., J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel and D. Hassabis. 2017. Mastering the game of Go without human knowledge. *Nature* 550(7676): 354–359. https://doi.org/10.1038/nature24270

Silver, D., S. Singh, D. Precup and R.S. Sutton. 2021. Reward is enough. *Artificial Intelligence* 299: 103535. https://doi.org/10.1016/j.artint.2021.103535

Storm, H., K. Baylis and T. Heckelei. (2020). Machine learning in agricultural and applied economics. *European Review of Agricultural Economics* 47(3): 849–892. https://doi.org/10.1093/erae/jbz033

Sutton, R.S. and A.G. Barto. 2018. *Reinforcement learning: An introduction*. MIT press, Cambridge, MA, USA.

Turgut, Y. and C.E. Bozdag. 2023. A framework proposal for machine learning–driven agent-based models through a case study analysis. *Simulation Modelling Practice and Theory* 123. https://doi.org/10.1016/j.simpat.2022.102707

Uthes, S., A. Piorr, P. Zander, J. Bieńkowski, F. Ungaro, T. Dalgaard, M. Stolze, H. Moschitz, C. Schader, K. Happe, A. Sahrbacher, M. Damgaard, V. Toussaint, C. Sattler, F.J. Reinhardt, C. Kjeldsen, L. Casini and K. Müller. 2011. Regional impacts of abolishing direct payments: an integrated analysis in four European regions. *Agricultural Systems* 104(2): 110–121. https://doi.org/10.1016/j.agsy.2010.07.003

van der Hoog, S. 2017. Deep learning in (and of) agent-based models: A prospectus. Bielefeld Working Papers in Economics and Management No. 02–2016. Department of Business Administration and Economics, Bielefeld University, Bielefeld. https://doi.org/10.4119/unibi/2900219

van der Hoog, S. 2019. Surrogate modelling in (and of) agent-based models: a prospectus. *Computational Economics* 53(3): 1245–1263. https://doi.org/10.1007/s10614-018-9802-0

Vargas–Pérez, V.A., P. Mesejo, M. Chica and O. Cordón. 2023. Deep reinforcement learning in agent-based simulations for optimal media planning. *Information Fusion* 91: 644–664. https://doi.org/10.1016/j.inffus.2022.10.029

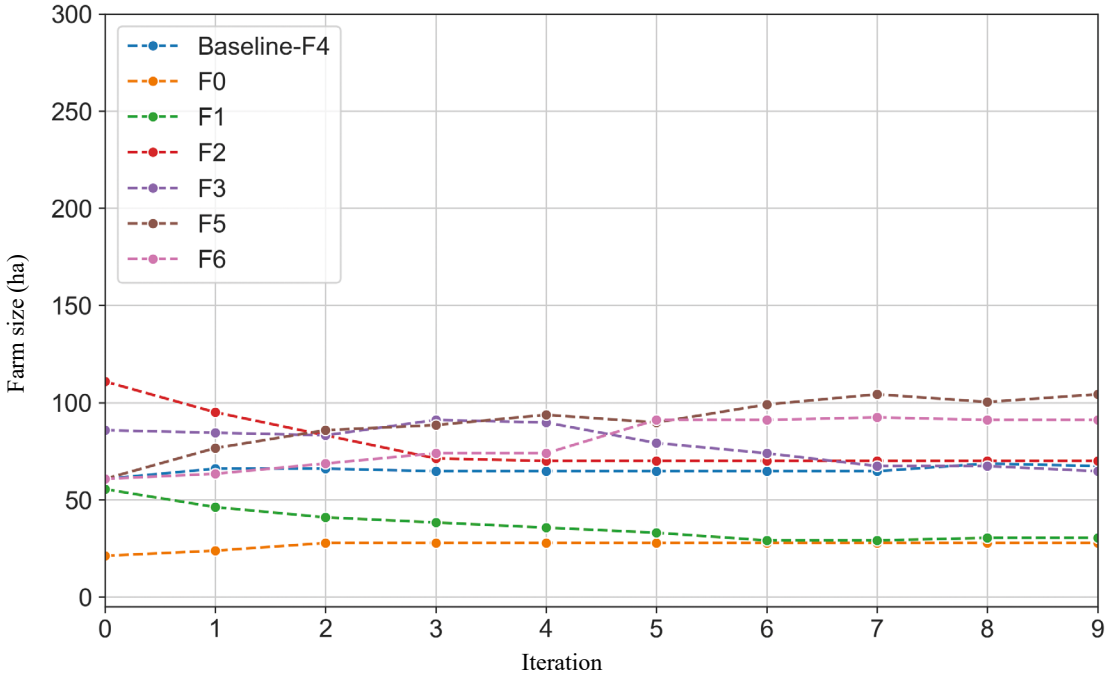
Zhang, W., A. Valencia and N.B. Chang. 2023. Synergistic integration between machine learning and agent-based modeling: a multidisciplinary review. *IEEE Transactions on Neural Networks and Learning Systems* 34(5): 2170–2190. https://doi.org/10.1109/TNNLS.2021.3106777

Appendix

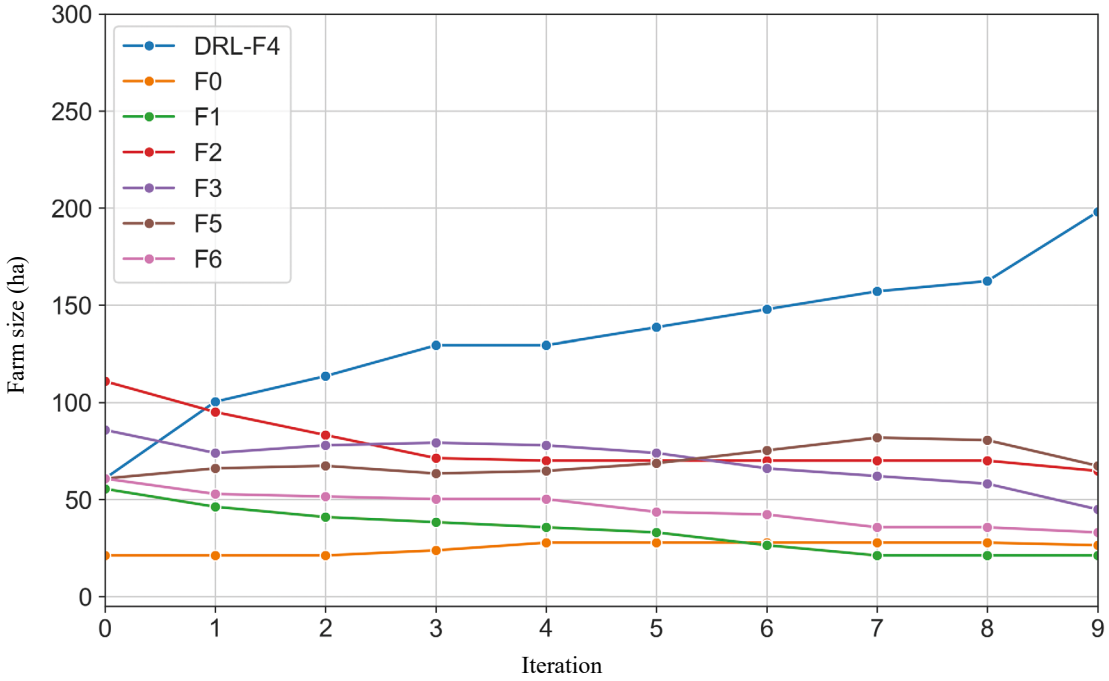
Table A1. Description of farm at initialization (iteration 0).

	Farm name						
	F0	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**
Farm type	Crop farm	Crop-farm	Crop farm	Dairy farm	Mixed farm	Mixed farm	Mixed farm
Land (ha)							
Owned land	10	22	14	32	24	24	31
Rented land	13	35	97	57	40	40	31
Selected indicators							
Farm age	43	52	44	55	44	64	64
Variable cost coefficient[a]	0.89	0.99	0.89	1.07	1.07	1.00	0.95
Land assets (× €1000)	166.06	373.31	616.04	488.11	368.93	368.93	510.23
No of investments[b]	1	1	4	6	6	6	6
Type of investment	Machine6	Machine5	Extcattle4	Dairy4	Pigs3	Pigs3	Pigs4
			Machine4	Robot2	Intcattle2	Intcattle2	Intcattle2
			Machine5	Vealer3	Dairy2	Dairy2	Dairy1
				Machine5	Parlour2	Parlour2	Parlour1
				Machine6	Vealer3	Vealer3	Machine5
				Machine5	Machine5	Machine5	Machine7
					Machine7	Machine7	
Liquidity (× €1,000)	−10.51	−57.34	−102.47	533.45	151.47	111.58	122.45
Equity capital (× €1000)	176.71	328.65	544.13	1181.45	782.33	782.33	782.33

This is reflecting the farms' managerial ability by manipulating the variable costs. There are 47 different types of investments with different capacities, lifetimes and costs. The agents are free to dispose, buy or even switch to different type of machinery. There are 43 different crop and livestock production activities.

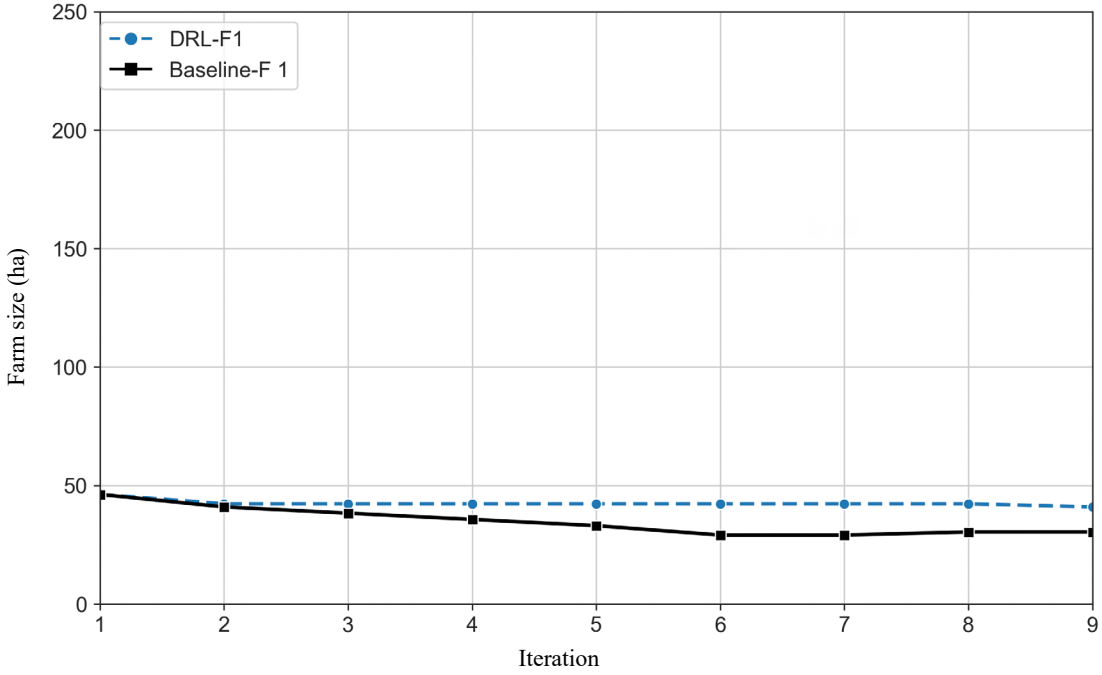


(a) Baseline

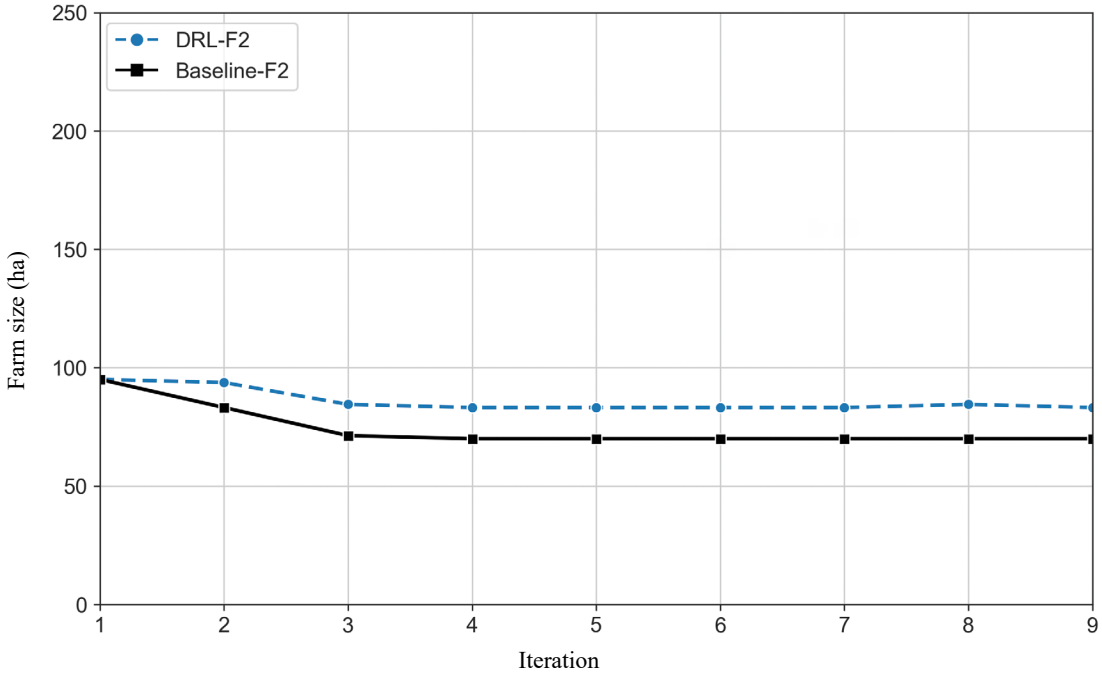


(b) DRL

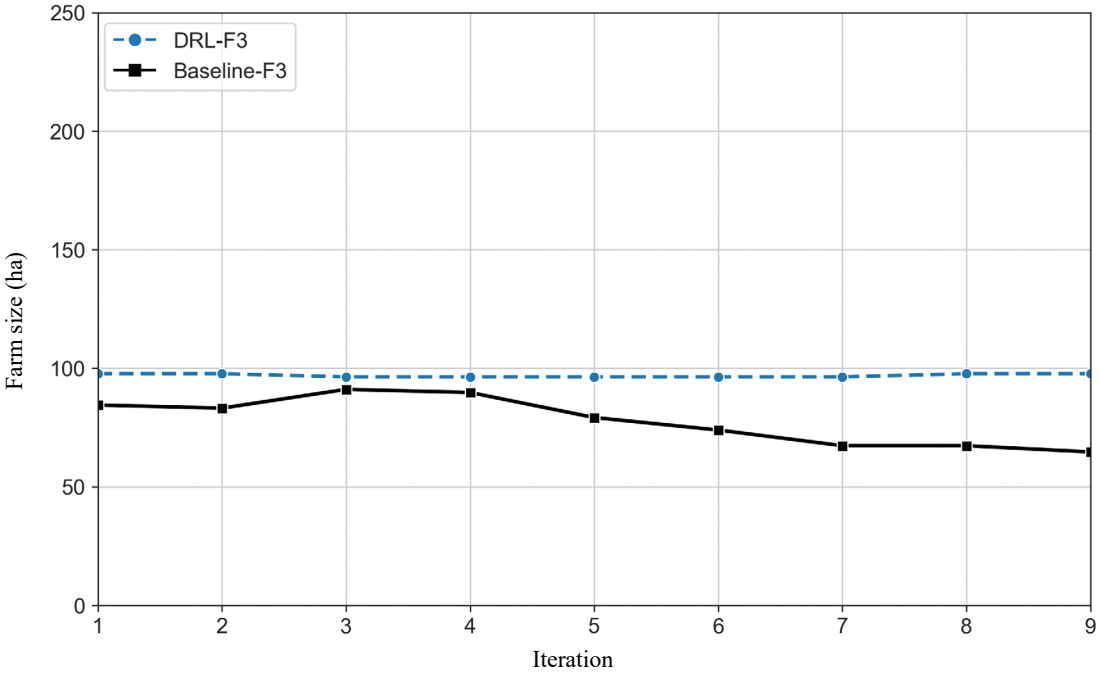
Figure A1. Effect of learning on the farm sizes of all the farms in the region



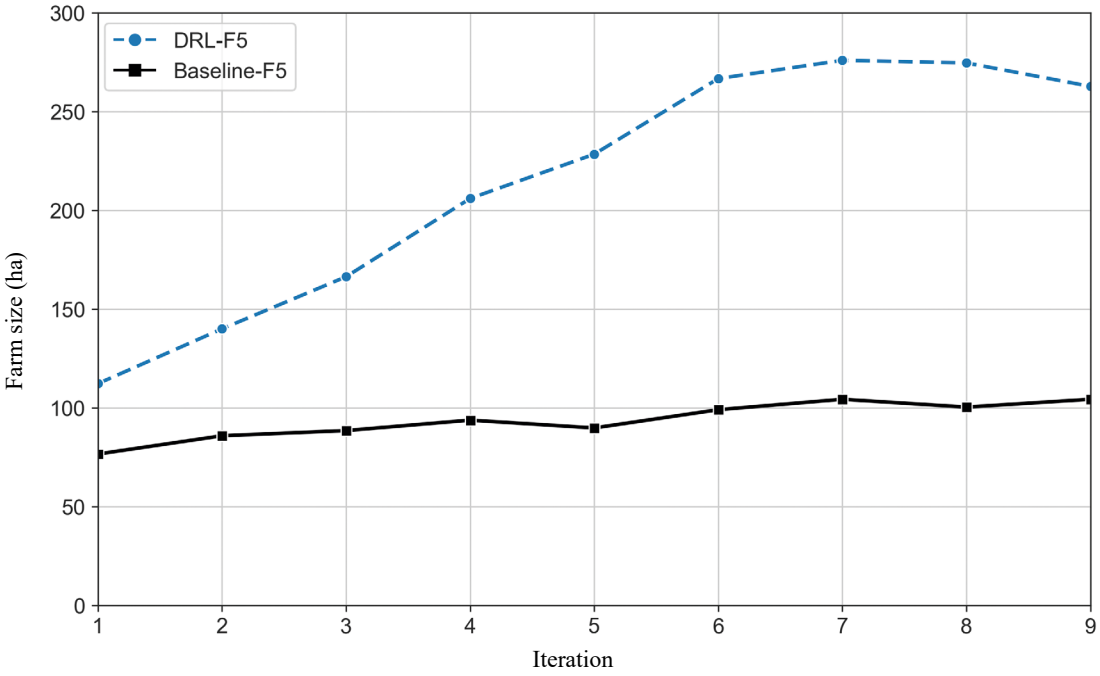
(a) Farm 1



(b) Farm 2

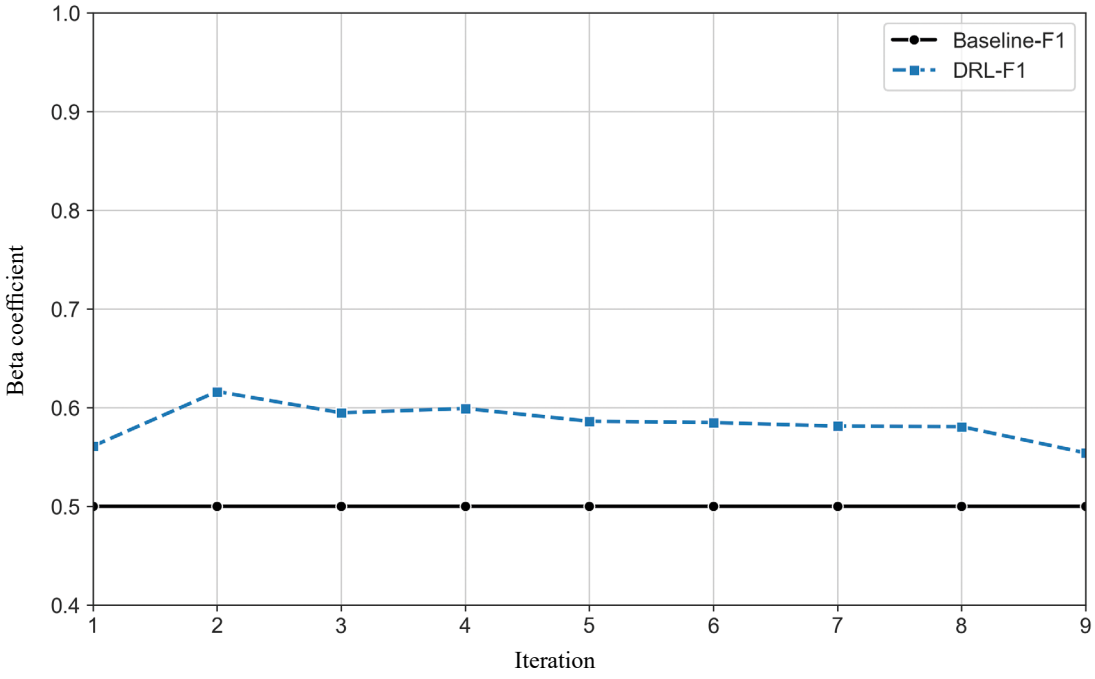


(c) Farm 3

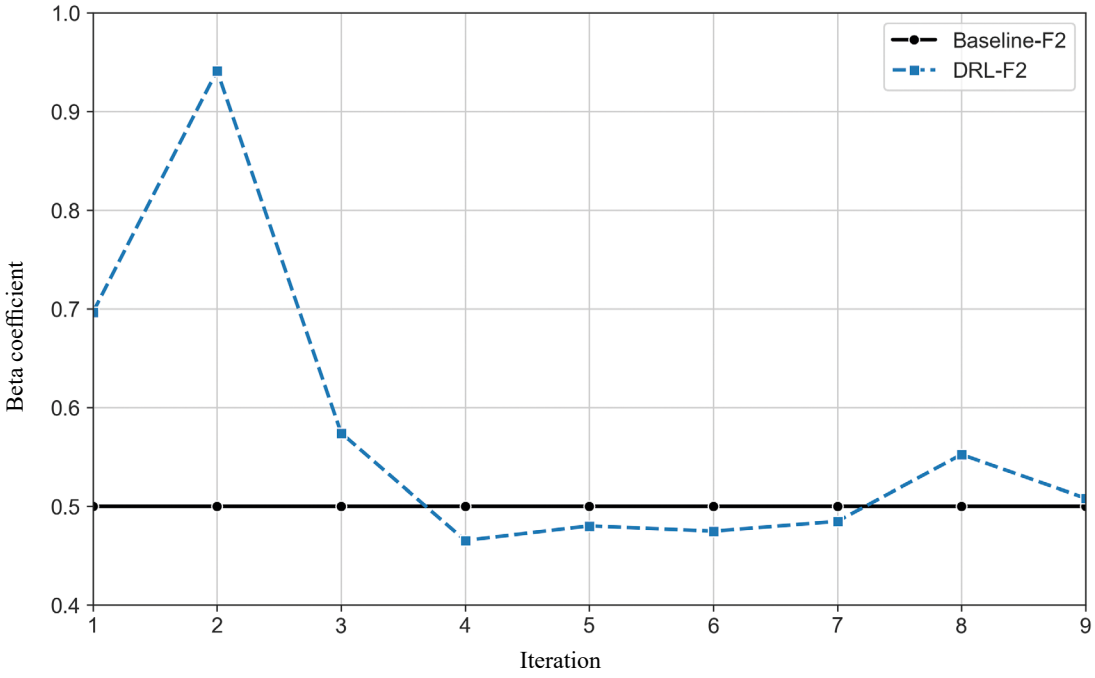


(d) Farm 5

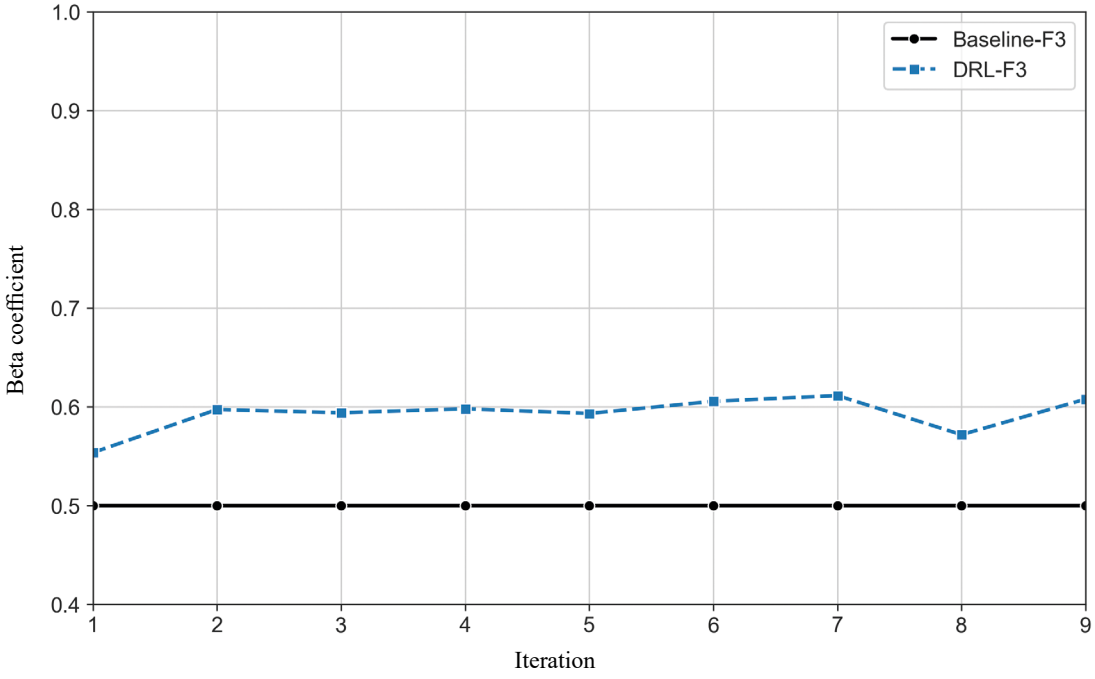
Figure A2. Evolution of farm size for the remaining farms.



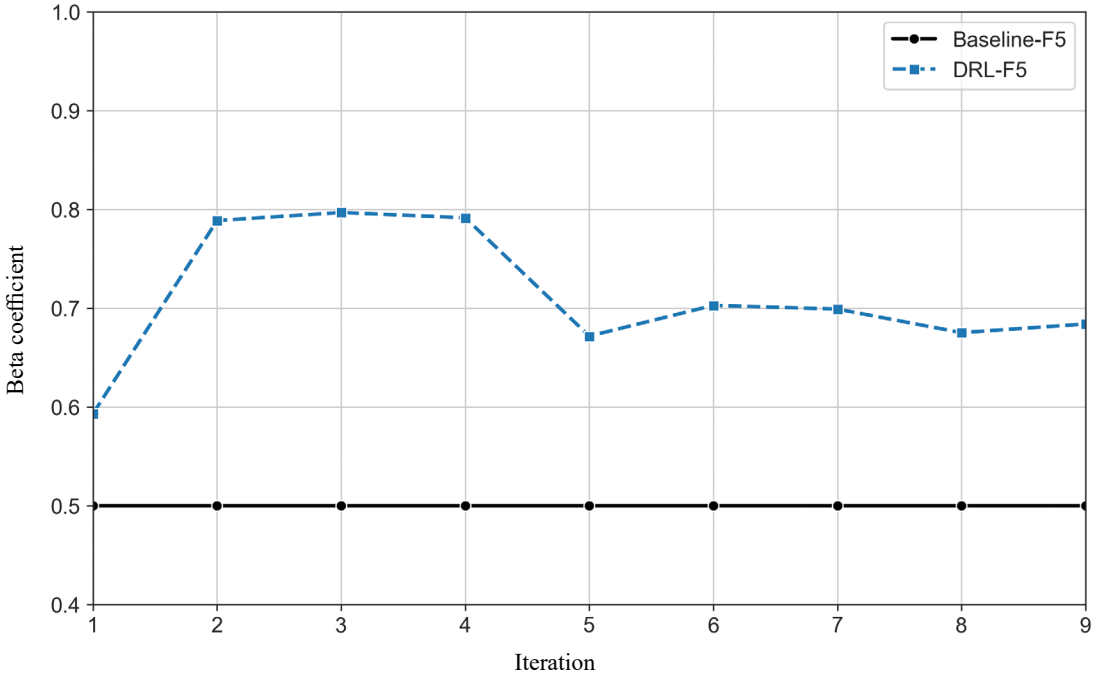
(a) Farm 1



(b) Farm 2



(c) Farm 3



(d) Farm 5

Figure A3. Best bidding strategy for the remaining farms.

Table A2. Comparison between fixed, optimal fixed and flexible rental factor (***β***).

Farm agent	Cumulative reward		
	Baseline (β=0.5)	**Optimal fixed β (optimal β in parentheses)**	**DRL (flexible β)***
F0	2 288 694	2 291 395 (0.51)	2 287 647
F1	3 932 183	4 023 643 (0.59)	4 022 778
F2	6 850 762	6 896 696 (0.44)	6 976 853
F3	12 669 940	12 833 917 (0.59)	12 868 645
F4	7 387 073	7 840 896 (0.62)	7 925 292
F5	7 889 960	8 211 464 (0.77)	8 613 919
F6	8 354 720	9 025 105 (0.75)	9 120 867

* See Figure A2.

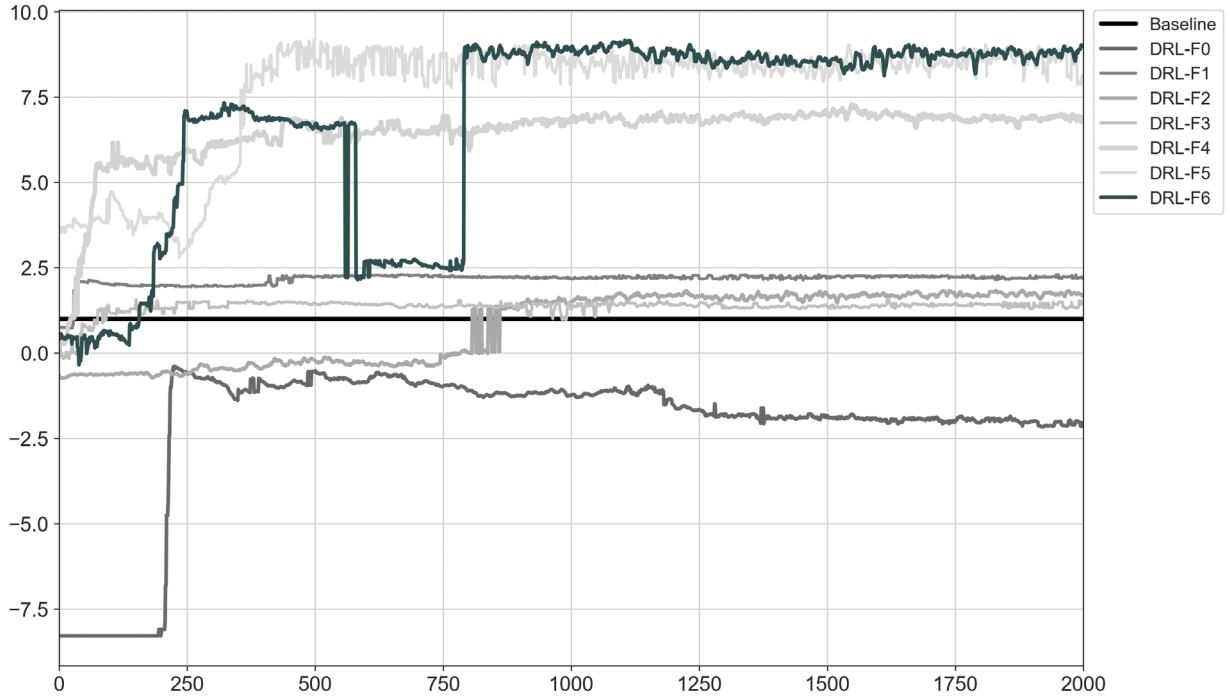


Figure A4. Relative change in cumulative reward in DRL compared to the Baseline for all farms. *See Figure 10, Results.