Article

Deep Q-Learning and Preference Based Multi-Agent System for Sustainable Agricultural Market

María E. Pérez-Pons [1] , **Ricardo S. Alonso** [1,2,*] , **Oscar García** [1] , **Goreti Marreiros** [3]
and Juan Manuel Corchado [1,2,4,5]

1   BISITE Research Group, University of Salamanca, Edificio Multiusos I+D+i, Calle Espejo 2,
    37007 Salamanca, Spain; eugenia.perez@usal.es (M.E.P.-P.); oscgar@usal.es (O.G.);
    corchado@usal.es (J.M.C.)
2   Air Institute, IoT Digital Innovation Hub, 37188 Salamanca, Spain
3   GECAD—Research Group on Intelligent Engineering and Computing for Advanced Innovation and
    Development, Institute of Engineering, Polytechnic of Porto, 4200-072 Porto, Portugal; mgt@isep.ipp.pt
4   Department of Electronics, Information and Communication, Faculty of Engineering,
    Osaka Institute of Technology, Osaka 535-8585, Japan
5   Pusat Komputeran dan Informatik, Universiti Malaysia Kelantan, Bachok 16300, Kelantan, Malaysia
*   Correspondence: ralorin@usal.es

Abstract: Yearly population growth will lead to a significant increase in agricultural production in the coming years. Twenty-first century agricultural producers will be facing the challenge of achieving food security and efficiency. This must be achieved while ensuring sustainable agricultural systems and overcoming the problems posed by climate change, depletion of water resources, and the potential for increased erosion and loss of productivity due to extreme weather conditions. Those environmental consequences will directly affect the price setting process. In view of the price oscillations and the lack of transparent information for buyers, a multi-agent system (MAS) is presented in this article. It supports the making of decisions in the purchase of sustainable agricultural products. The proposed MAS consists of a system that supports decision-making when choosing a supplier on the basis of certain preference-based parameters aimed at measuring the sustainability of a supplier and a deep Q-learning agent for agricultural future market price forecast. Therefore, different agri-environmental indicators (AEIs) have been considered, as well as the use of edge computing technologies to reduce costs of data transfer to the cloud. The presented MAS combines price setting optimizations and user preferences in regards to accessing, filtering, and integrating information. The agents filter and fuse information relevant to a user according to supplier attributes and a dynamic environment. The results presented in this paper allow a user to choose the supplier that best suits their preferences as well as to gain insight on agricultural future markets price oscillations through a deep Q-learning agent.

Keywords: multi-agent systems; decision support systems; sustainable agriculture; deep Q-learning; IoT; edge computing

1. Introduction

Environmental factors such as the consequences of climate change [1] directly affect producers in the agricultural price setting process [2]. Agricultural trade is expected to grow at roughly half the rate of the past ten years [3]. Nevertheless, for most commodities, the share of total production that is being traded on global markets will remain relatively constant. Technology will also be more important in guaranteeing global food security, due to natural resource constraints in many countries [3]. In this regard, the concept of bioeconomy is gaining importance within the European Union objectives for the 2030 Agenda and its Sustainable Development Goals, and new indicators to measure the performance of bioeconomy sectors are emerging [4]. Globalization and the possibility of

exporting and importing, as well as competing on a larger market, have led to a paradigm in which agricultural products are a financial indicator of the economy and take part in the prices of future markets. Agricultural commodity prices also tend to correlate with trends on energy markets, with oil prices experiencing unusual rises and falls in recent years [5]. Although agricultural inputs, production, storage, and transportation have long been influenced by energy prices, the rapid growth of the biofuels sector has created new types of relationships between agricultural and energy markets [5]. Moreover, in periods of financial crises, the volatility of the agricultural market has been very high [6,7]. Those fluctuations on commodity futures markets [8] contribute to the agricultural price oscillations for the farmers. However, not all agricultural products are affected by the same factors, the agricultural market is very diverse in terms of product attributes and particularities which are subject to external factors. For instance, corn futures prices may be determined by certain factors (such as location, transportation costs, contamination, warehouses, weather conditions, etc.) in comparison to wheat prices, that are less sensitive to weather. The changes in climate are something that has been more accentuated in recent years due to the greenhouse effect [9]. Analysis of greenhouse emissions as well as limited resource expenditures or energy consumption is becoming more widespread in precision agriculture (PA) [10] and smart agriculture (SA) [11].

PA is a term that has been coined in recent years. It refers to the concept of using new technologies to increase the yield and profitability of crops while reducing the resources needed for cultivation [12,13]. Thanks to technological advances, farms in developed and developing countries can benefit from the application of low-cost technologies. In this regard, the Internet of Things (IoT) and, more specifically, the Industrial Internet of Things (IIoT), is presented as a key enabling technology for implementing and monitoring resource management solutions in various scenarios in Industry 4.0, including smart agriculture environments [11]. The monitoring of all these values can be carried out with sensors, using new paradigms such as edge computing which enable monitoring while reducing the cost of data analysis in the cloud [14,15] increase the efficiency of agricultural processes [16]. The application of multi-agent systems to monitor agricultural processes is common, especially to achieve the efficient use of land in terms of investment and production [17], to manage resources [18], to increase the efficiency of irrigation systems [19,20], to optimize energy use [21], or predict the prices of agricultural products [22]. Moreover, in the field of selecting suppliers, Valluri and Croson [23] conducted a research towards best supply selection through a game theory approach with agents, in a scenario where reward and punishment were complicated by incomplete information. Over the last years, the number of processes oriented to sustainability objectives [24] and to developed tools [25] is growing. For instance, it has become more popular to integrate environmental, economic, and social attributes when selecting a supplier and sourcing process [26–28]. There are many differences in price setting in terms of requirements and the most important attributes for which each model can be built. Most models rely on historical data; however, others focus on the buyer, in which there are sales quotas, decision history polynomials, probabilities, and regressions of the potential prices that might be accepted [29]. In the literature, other approaches have been found, such as agent-based modeling with reinforcement learning conditioned by inventory [30,31]. When product pricing strategies are connected to estimations of price allocation and real-time reporting is taken into account, the relationship between data and price allocation parameters can be modeled dynamically, as demonstrated in the multi-agent supply chain [32]. In the context of a market, goods are exchanged and there are two main players. One of the basic rules in a price system is that goods can be exchanged according to the relative prices of the goods in question. Twenty years ago, Wellman and Wurman [33] already developed the first market-based multi-agent paradigm at a theoretical level.

The above examples demonstrate that multi-agent systems are applied in cases where a single agent is not capable of carrying out all the processes, as several agents and objectives interact in real-time [34]. Although currently there are different MAS focused on price

forecast [35] or choice of suppliers [26–28], to this day, there is still no MAS in which a buyer can have information on the sustainability of a supplier or a product by monitoring the greenhouse emissions involved in production, the use of pesticides or the consumption of natural resources and having access to information on prices on the futures market. Considering that these needs cannot be fulfilled with a single agent, a MAS has been built [36]. The proposed system consists of MAS that helps choose a supplier through decision-making based on certain attributes. These attributes measure the sustainability of a supplier and also have prices of agricultural futures markets. The different attributes that have been considered are agri-environmental indicators (AEIs) [37]. These parameters consist of water consumption, greenhouse emissions, energy consumption, as well as the use of edge computing technologies that represent an improvement in terms of data transfer costs to the cloud. More concisely, the MAS consists of a preference-based multi-objective optimization problem that is open to real implementation, and therefore each user can fix the desired threshold and input requirements. Although not all products are comparable at the world level, the developed MAS is intended to give an indication of the most recent price developments. Futures markets are an important source of price information for farmers, but only a small percentage of farmers directly trade futures. The availability of high-frequency (intraday) data can help market participants make quicker decisions compared to low-frequency data, such as daily or monthly data. Having high-frequency data allows to better forecast the stock prices [38] so that farmers can sell according to trends in agricultural futures market real prices without having to wait to learn of the effects of selling products.

The rest of this paper is structured as follows. Section 2 consists of a revision of the state of the art of technologies involved in the system. Section 2.1 introduces the edge computing technologies as they play a significant role in monitoring and cost reduction in agriculture and identifying the most important trends in the application of those paradigms in smart farming scenarios. Then, in Section 2.2, state-of-the-art MAS are described, specifically those designated for agriculture scenarios, and finally, in Section 2.3 the deep Q-learning concept is introduced. Section 3 describes the MAS architecture and the data that have been used to conduct the experiment. Section 4 describes the experiments that have been conducted and the results. Finally, Section 5 discusses the solution, implementation fields, conclusions, and future work.

2. Related Work

This section reviews the state-of-the-art of the three main topics that are directly related to the case study: the edge computing paradigm, the MAS, and the deep Q-learning algorithm. First, the state-of-the-art of edge computing (EC) paradigm is presented. EC is a paradigm that enables reducing data transmission costs to the cloud, and in this case when conducting the analysis of consumption and greenhouse emissions. Second, the MAS and the multi-objective optimization processes are described to contextualize the conducted experiment in which different agents interact to identify the best supplier and price according to a given preference, and finally an introduction to Deep Q-learning which is the technique used for forecasting the agricultural future market prices.

2.1. Edge Computing

The increasing demand for food in terms of quality and quantity has increased the need for industrialization and intensification in the agricultural field [39]. Internet of things (IoT) is a very promising technology that offers many innovative solutions to modernize the agricultural sector [40]. IoT can be used in combination with other technologies such as cloud computing, big data, AI, or distributed ledger technologies (e.g., blockchain) to implement solutions that improve the traceability and productivity of industrial processes [41]. However, when trying to transmit data to the cloud, several challenges arise regarding the privacy of the data, power consumption, or costs associated with the use of cloud services [15]. In this regard, service providers charge fees according to the amount of data

that is transferred, stored, and processed in the cloud [42]. By using EC technologies, it is possible to reduce the traffic between the IoT layer and the cloud [14]. EC allows for the execution of machine learning models at the edge of the network, reducing the response time and providing a certain level of service even if the communication with the cloud is interrupted. This is commonplace in scenarios where Internet connectivity is limited (for example, rural agricultural environments) [15]. The EC paradigm has been also used in different studies in which results show that including the costs of edge and non-edge data transfer has an impact on the efficiency [16].

2.2. Multi-Agent Systems

Preference-based multi-objective optimization has had an increasing interest in research and academia in the last years [43]. Agents can be defined as intelligent entities with social skills (communication, collaboration, interaction, negotiation, intelligence, coordination, competence) that encapsulate a functionality to solve a problem [34,44]. When two or more agents are able to work together in order to solve a common problem, they form a MAS [34]. MAS are systems that integrate a set of agents that interact, communicate, and coordinate to achieve the established objectives [45]. MAS are designed to meet a set of objectives according to a set of rules and standards. The different designs of MAS have different nomenclatures depending on the methodology, nevertheless, they tend to include social, communicative, interactive, and normative aspects [46]. Each of these is described below.

1. Social aspects refer to the description of the set of roles, groups (role associations), and the relationship between them. Regarding the existing relationships between roles and groups (recursively), some authors have defined a set of social structures that allow to model the interactions between members. Among the main structures, the following stand out: hierarchies, coalitions, teams, congregations, societies, federations, markets, matrices, and composite organizations. Some studies have simply defined possible relationships between members [46] such as dependency, hierarchy, use, etc.
2. The communication aspects refer to the means that makes the exchange of information possible. That is, a knowledge representation language (usually represented by an ontology) and a communication language. The communication sequence between two agents is called illocution [47], communication act [46], or link [48].
3. Interaction aspects refer to how roles collaborate to achieve common goals. There might be objectives that cannot be achieved individually, and that require the combination of several agents for achievement, and it is necessary to describe an interaction structure that allows to articulate or regulate the achievement of individual sub-objectives that in turn make the achievement of higher-level objectives possible [46].
4. Normative aspects: note that this is one of the main pillars of organizational MAS [49]. Norms (or institutional patterns make it possible to establish a relationship of trust between the members of an organization, as they limit the free will of individual agents [50].

In addition to the concepts that have just been presented (role, organization, norms, and social structures), organizational MAS routinely include another key concept: Environment. Agent theory traditionally conceives the agent as an entity that plans its actions on the basis of its perception of the environment. However, the increasing complexity of the environment itself in the context of open systems (dynamic, heterogeneous, and unpredictable) can not only make the MAS unpredictable, but also difficult to interact with [34,44].

Moreover, it is important to understand the effect of ubiquitous automated agents on the performance of economic systems. With a special emphasis on being able to achieve, at the computational level, the capacity for some agents to reason about the reasoning of other agents and of humans who would also be at stake. Moreover, these agents will adopt a game theory vision [51], where each agent will act according to the behaviour of the other agent (in the best and most rational manner for both agents). Game

theory consists of a mathematical theory that studies interactions among self-interested agents [52]. The traditional game theory was revised and applied in biology by the authors of [53], in which the authors determined the concept of Evolutionary Stable Strategies (ESS). In these strategies, there were not only two players in a complete information situation and that was a condition of Nash's equilibrium [54]. Therefore, it would lead to equilibrium situations that are part of the traditional economy. The quality of an AI design is determined through the degree to which the agent's actions achieve specific objectives, subject to observed perceptions. If we express objectives in terms of preference over results and perceive both perception and action within the framework of decision-making under uncertainties, then the position of the AI agent is fully in line with the standard economic paradigm of rational choice. Consequently, the task of the AI developers is to build rational agents, or agents that are as rational as possible, given the limits of their computational resources [55]. At the multi-agent level, a developer cannot directly program the behaviour of AI, but instead determines the rules and incentives that will regulate the interactions between AI. The authors of [56] propose a multi-agent system to simulate group decision-making processes, where agents are designed with emotional properties and reason using incomplete information.

In real-world scenarios, there are multiple applications and situations within a given market, different agents have to make decisions with incomplete information. Methods such as the game theory for portfolio optimization can be used in these cases, regardless of the product in question [57]. The application of computing techniques for the optimal product or supplier portfolio, from the application of machine learning algorithms [58], to genetic algorithms for product optimization [59], neural networks [60], deep neural networks [61], and reinforcement learning [62].

Multi-Objective Optimization Problem

A multi-objective optimization problem can be defined as the following Equation (1), where $f(x)$ is the k-dimensional objective vector.

$$\textbf{Max} f(x) = (f_1(x), f_2(x), \dots f_k(x)) \tag{1}$$

The multi-attribute utility function is used to represent the preferences of a user over packages of goods, under conditions of certainty about the results of any potential choice. Van Calker et al. [63] presented a model of the sustainability multi-attribute function for evaluating sustainability in different farming systems.

Preferences can be characterized by utility functions, where the information regarding preference is implicitly involved in the function, enabling the ranking of solutions. Utility functions assign different weights to given attributes. The utility function for the buyer agent in this case would be the following, in which product A (which could be Corn sold at X price and Greenhouse contamination of Y points), is preferred over product B only if the expectation of the function U is higher under A than under B, as shown in Equation (2).

$$E_A[u(x_1, \dots, x_n)] > E_B[u(x_1, \dots, x_n)] \tag{2}$$

2.3. Deep Q-Learning Algorithm

Reinforcement learning (RL) consists in an agent interacting with the environment, learning an optimal policy, by trial and error, for sequential decision-making problems [64]. The standard RL consists of an agent interacting with an environment, which can be modeled as a Markov decision process (MDP).

The Q-learning algorithm [65] is one of the best-known, model-free techniques in RL and has numerous evolutions and variants [66]. Q-learning, is a model-free off-policy RL method, which consists of agents whose objective is to reach the state-action-value of a function $Q = (s, a)$ by interacting in a given environment. As the agent explores the environment, Q returns an increasingly accurate approximation of the expected value of an action a, given a state s of the expected value of an action a, given a state s. That is, the

function Q is progressively updated. Q-learning [67] can be defined as a way for agents to learn how to act optimally in controlled markovian domains, which means the future depends only on the current state and action, but not on the past. It is formulated as an MDP which can be defined by the 5-tuple (s, a, p, r, γ), where s is the state, a is the action, p is the transition probability, r is the reward function, and γ is the discount factor.

Deep learning (DL) has accelerated progress in RL, with the use of deep learning algorithms within RL defining the field of deep reinforcement learning (DRL) [66]. DL allows RL to be extended to previously intractable decision-making problems, i.e., environments with a high number of dimensional states and action spaces. As a neural network is a universal functional approximation, it can be used as a substitute for the Q-table. In the learning process, DL optimizes the weights, θ, to minimize the error estimated by the loss function. The error or loss is measured as the difference between the predicted result and the actual result. The deep Q-network (DQN) was first introduced by [68] and then [69] introduced additional techniques, such as DQL. The base algorithm for DQN is value-based RL, which is a method that approximates an action value (i.e., a Q-value) in each state. An algorithm based on Q-learning that approximates the Q-function using DNN is the basis of DQN [69]. To prevent DNN from learning only through the experience of a specific situation, experience replay has been introduced to sample a general experience batch from memory [69]. In reinforcement learning, the temporal difference (TD) target function is always unknown. Before an agent takes an action, the Q-value can be defined as $Q(s, a)$, and after the action is taken the new state is $R(s, a) + \gamma max_{a'} Q(s', a')$, so the temporal difference is defined in Equation (3).

$$T(a, s) = R(s, a) + \gamma \times max_{a'} Q(s', a') - Q_{t-1}(s, a) \tag{3}$$

The value function is approximated by a neural network $Q(s, a; \theta)$ with a parameter, θ where the parameter is learned by minimizing the TD loss. Thus, the loss function turns out to be

$$Loss = Q^*(s_t, a_t) - Q(s_t, a_t) \tag{4}$$

The key idea of DQN is to learn an approximation of the optimal value function Q, which conforms to the Bellman optimality equation [70]. In the DQN algorithm, the Q corresponds to the function that represents the expected rewards for a given action in a given state. DQN refines the policy with respect to action values by the max operator [71]. One way to minimize the loss function is by the gradient descent method [72]. In this method, the policy $Q(s, a)$ is updated on the basis of the current reward and the maximum value of the expected future rewards. In the DQN, the learn function can be described as in Algorithm 1, where ϵ is the learning rate and π is the optimal policy.

Algorithm 1: Algorithm adapted from the work in [71].

Result: Q
initialization; $Q : X \times A \to R$
while *Q has not converged* **do**
    Start in states $s \in X$
    while *s is not terminal* **do**
        calculate π according to Q and exploration strategy
        $a \leftarrow \pi(s)$
        $r \leftarrow R(s, a)$
        $s' \leftarrow T(s, a)$
        $Q(s', a) \leftarrow (1 - \alpha) \times Q(s, a) + \alpha \times (r + \gamma \times max_{a'} Q(s', a'))$
        $s \leftarrow s'$
    end
end

---

3. Case Study

This section describes the MAS that has been designed and how it has been conducted. This section has two main parts: first, the architecture of the MAS is defined in Section 3.1, and second, the different types of data used to perform the experiment are described in Section 3.2.

3.1. Architecture

The architecture that has been designed for this case study is represented in Figure 1. A MAS architecture has been used as a MAS can be defined as a collection of, possibly heterogeneous, computational entities, having their own problem-solving capabilities and which are able to interact in order to reach an overall goal. In the case of this investigation, each agent has a role [73]. As shown in Figure 1, there are three agents that communicate to provide the best option according to user preferences; it is a closely collaborating agent system in which every agent has its own specialized capabilities and knowledge, and no single agent has full knowledge of the world. The main functions include agricultural futures market information retrieval, deep Q-learning for lowest price setting, and preference based weight system for users' preferences. The architecture designed for the MAS has been implemented with SPADE library [74].
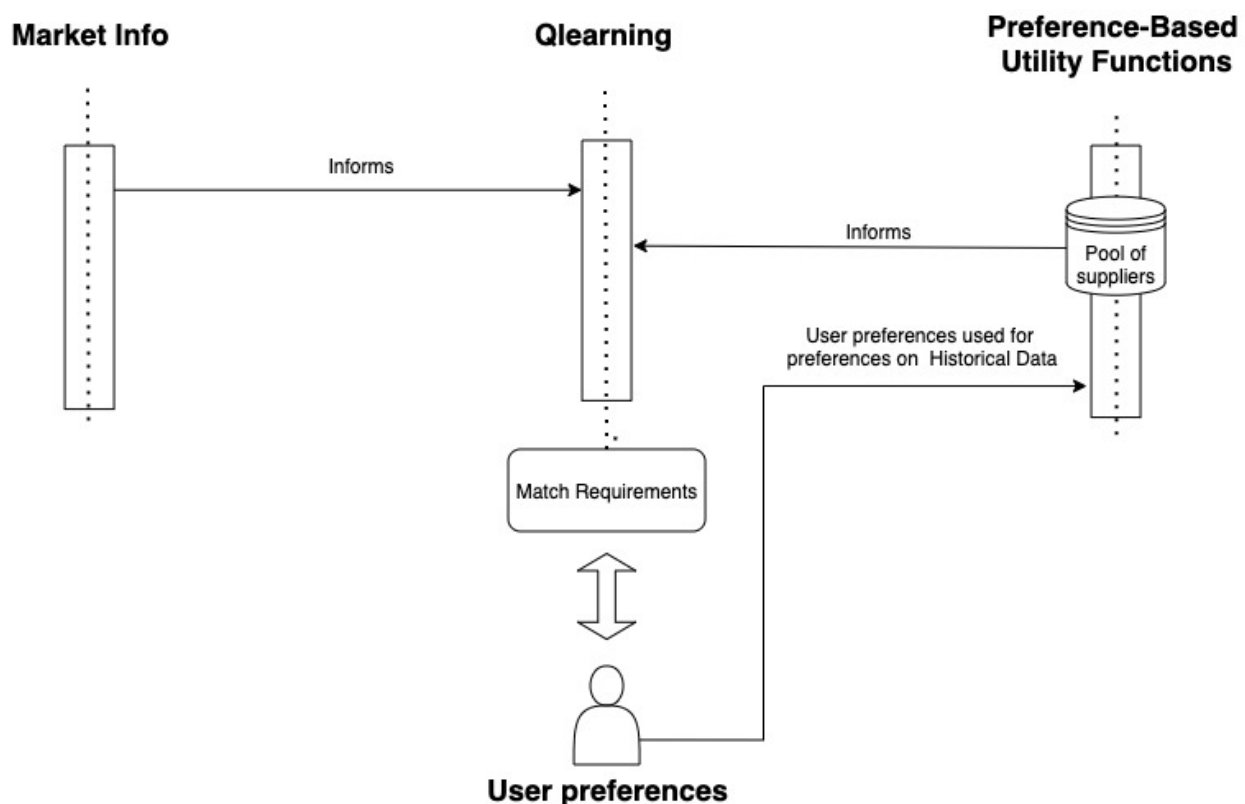

Figure 1. Proposed multi-agent system.

In this case, to test the MAS, one agent collects information from the period described in Section 3.2. Then, the preference-based utility functions retrieve information from the historical database and also the preferences of the user in terms of weights for each attribute. In the case of this study, as the approach is preference-oriented and there are some weights according to each attribute, the Equation is represented as in Equation (5). For the values of the other attributes, the equivalence is shown in Equation (6).

$$(w_2, \ldots, w_n)(w_2, \ldots, w_n) \tag{5}$$

$$(y_1, w) \sim < (x_1, w) : (z_1, w) > \tag{6}$$

Next, the deep Q-learning agent is responsible for modeling nonlinear trends of stock price time series, by predicting values and identifying the lowest prices.

3.2. Data

The data that have been chosen for testing the MAS can be modified and adapted to other products or markets. To implement the case study, information has been taken from a specific time-period, but as mentioned above the presented architecture is adaptable to any market. For the development of the model, two main datasets have been used; on the one hand real data from agricultural futures market and on the other hand a synthetic dataset for potential suppliers. The synthetic dataset has been built with different agri-environmental indicators (AEIs) [37] as main attributes for each supplier. AEIs track the integration of environmental concerns in the common agricultural policy (CAP) at EU, national and regional levels. The different attributes that have been considered are in Table 1 as follows:

Table 1. Supplier attributes.

Attribute	Description	Measure Units
Energy consumption	Energy consumed	kWh
Greenhouse	N_2O is produced mostly from excess nitrogen in soils; one way to suppress emissions of this gas is to apply fertilizer judiciously: adding just enough, at the right place and time, to meet crop demands, but avoiding excess amounts. This can reduce fertilizer costs for producers and reduce the amount of nitrogen lost through excess fertilizer application	N_2O
Water consumption	Amount of water consumed per year for the irrigation of crops	mL/ha
Edge Computing Techniques	Whether edge computer techniques are used to reduce the cost of using the cloud and make their sensors communications more robust and scalable with the cloud	Boolean (Yes/No)

In the case of the qualitative values, such as the edge computing attribute, the values were converted into the following values $[0, 1]$. The other data that are considered in the study are the data from the agricultural futures market. To evaluate the model and to see how it works, information has been collected from the corn agricultural futures market, which has the CZ symbol. The period for which the data has been collected is from 2017 till 2019, because the 2020 and 2021 values are very irregular due to the COVID-19 pandemic. The attributes that are usually collected from the stock market are the ones defined in Table 2; nevertheless, the most commonly used inputs for next-day stock price prediction in the literature are the stock index opening or closing prices [75], and those are the values have been gathered for the case study, as represented in Figure 2.

Table 2. Stock market price variation. Source: Stock market for the symbol CZ.

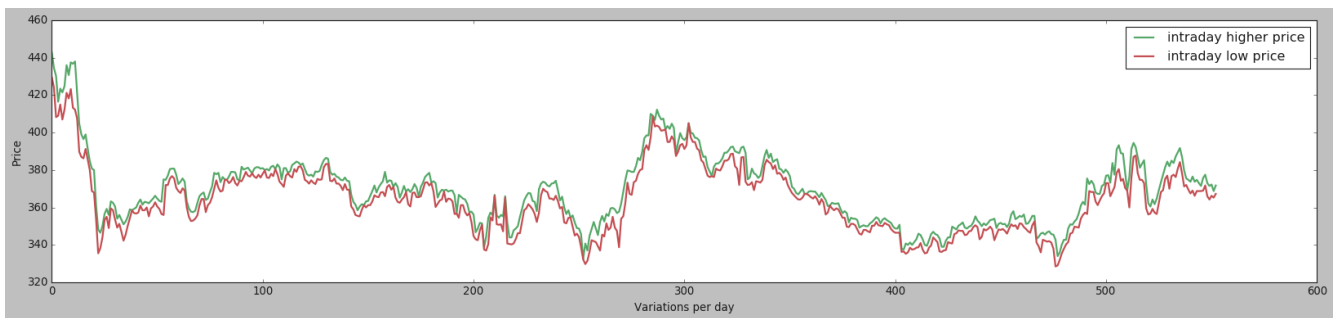| Attribute	Description
Date	The information regarding price variations is considered per day
Price	Stock price
Open	Stock price at the opening
High	Highest price within a concrete day
Low	Lowest price within a concrete day
Vol.	Number of stocks
Change %	Variation regarding previous date

Figure 2. CZ symbol price variation.

4. Results

The problem that the presented MAS overcomes is the difficulty of choosing a supplier according to sustainability parameters, as well as having transparent market information. The MAS presented helps potential buyers to identify a supplier according to their preferences in terms of sustainability as well as efficient use of resources. It also combines the selection of the supplier according to certain preferences, as well as seeing the actual price differences with the quotations of the products and being able to buy at the most optimal times. The designed MAS has two main results that are presented below. The first result is related to the supplier selection ranking, and the second one to price forecasting.

To achieve the first goal, different values have been assigned to the attributes, in terms of user preferences. To test the model, different cases have been tested with the following preferences as represented in Table 3.

Table 3. Case study combinations according to weights assigned [1,10] for the attributes of each supplier.

	Greenhouse	Water	Energy
Case 1	3	5	2
Case 2	0	8	2
Case 3	8	1	3

Then, a value between [0,10] is assigned to each supplier, where value is understood as the total sum of the weights of different attributes, assigned according to the attribute, and attribute importance. The outputs and the values of each of the parameters are represented in the Figures 3–5, in which three dimensions are represented; the value according to the different weights for each of the parameters.
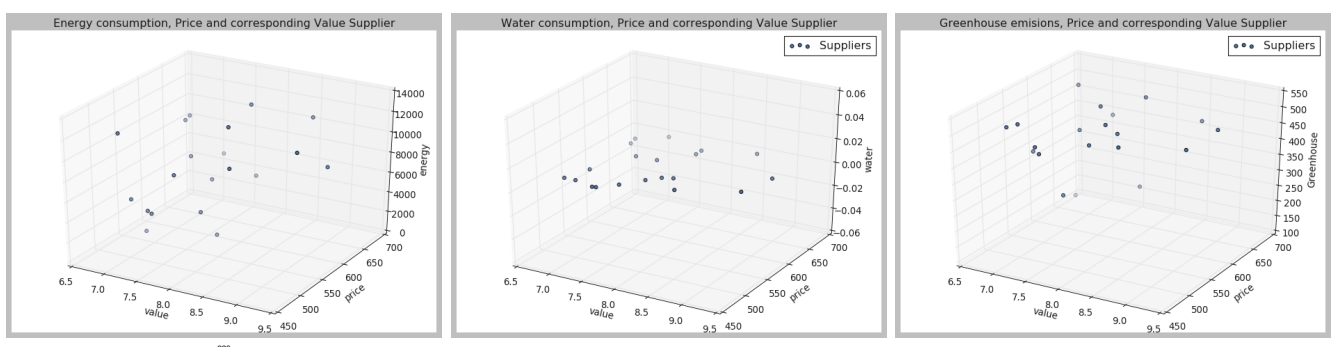

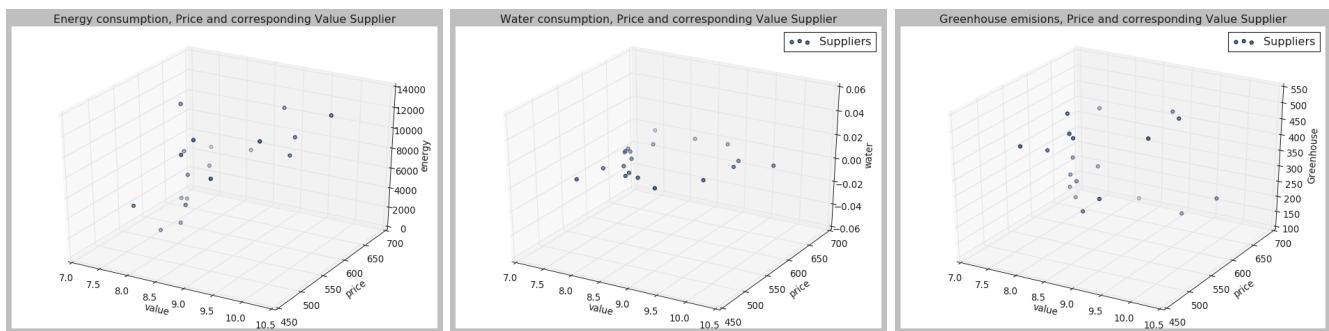**Figure 3.** Case 1, according to the preferences described in Table 3.

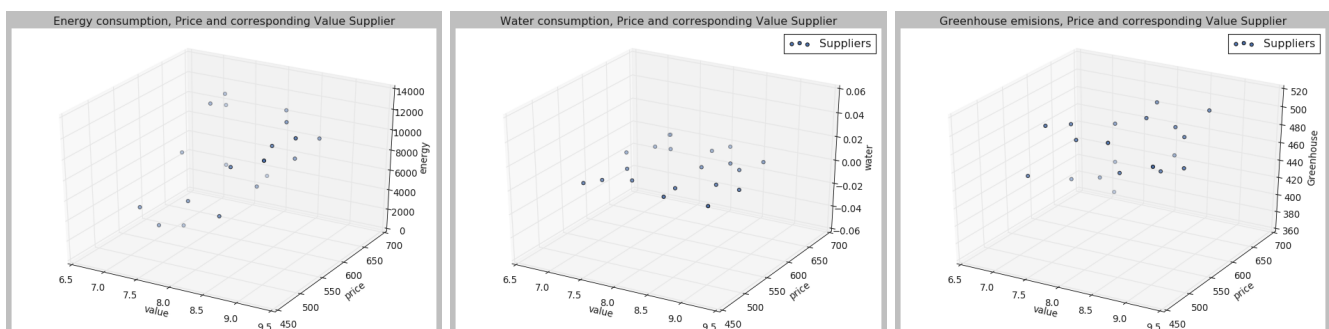**Figure 4.** Case 2, according to the preferences described in Table 3.

Figure 5. Case 3, according to the preferences described in Table 3.

As can be seen from Figures 3–5, the different suppliers' selection would vary according to the preferences set by the user. Once the different suppliers have been ranked, the results are sent to the Q-learning agent. The deep Q-learning agent's goal is to maximize the total amount of reward it receives. In this case, the agent's goal is to buy at the lowest price given a certain amount of money. Therefore, the user should fix an initial amount of money, which in this case has been 20,000. To build the agent, the different parameters have been fixed as shown in Table 4. Then, different hyperparameters have been changed to identify the best agent performance. The set hyperparameters are shown in Table 4, and then the other hyperparameters have been compared, as represented in Table 5.

Table 4. Fixed hyperparameters that have been used for the different case combinations in the deep Q-learning agent.

Parameter	Description	Value
γ	Maximizes the current reward	0.950
ϵ	Either taking random actions or using the trained actions	0.500
ϵ decay	The decrease over time in the use of the random and trained actions	0.999
Actions	The actions that can be taken by the agent, which is either selecting a buying price or do nothing	2
Replay memory size	Agent's experiences at each time step in a data set	1000

The Q-learning agent has one hidden layer with 256 neurons and the activation function is the rectified linear unit (ReLU) [76]. In the learning process, DL optimizes the weights, θ, to minimize the error estimated by the loss function. Therefore, the error or loss are measured as the difference between the predicted result and the actual result. The loss function for the different cases which are described in Table 5, is represented in Figure 6.

Table 5. Parameters used for each case comparison related to the different cases.

Case Number	Window Size	Batch Size
1	8	32
2	10	32
3	20	32
4	8	64
5	10	64
6	20	64

As can be seen in Figure 6, cases 3 and 4 are the ones achieving the best results in all Figure 6a–c, while, for instance, case 1 has the worst results in Figure 6a but then has better results in Figure 6b. Moreover, in Figures 7 and 8, all the marked values are placed in the real price variation line.

(**a**) 600 iterations

(**b**) 1200 iterations

(**c**) 2400 iterations
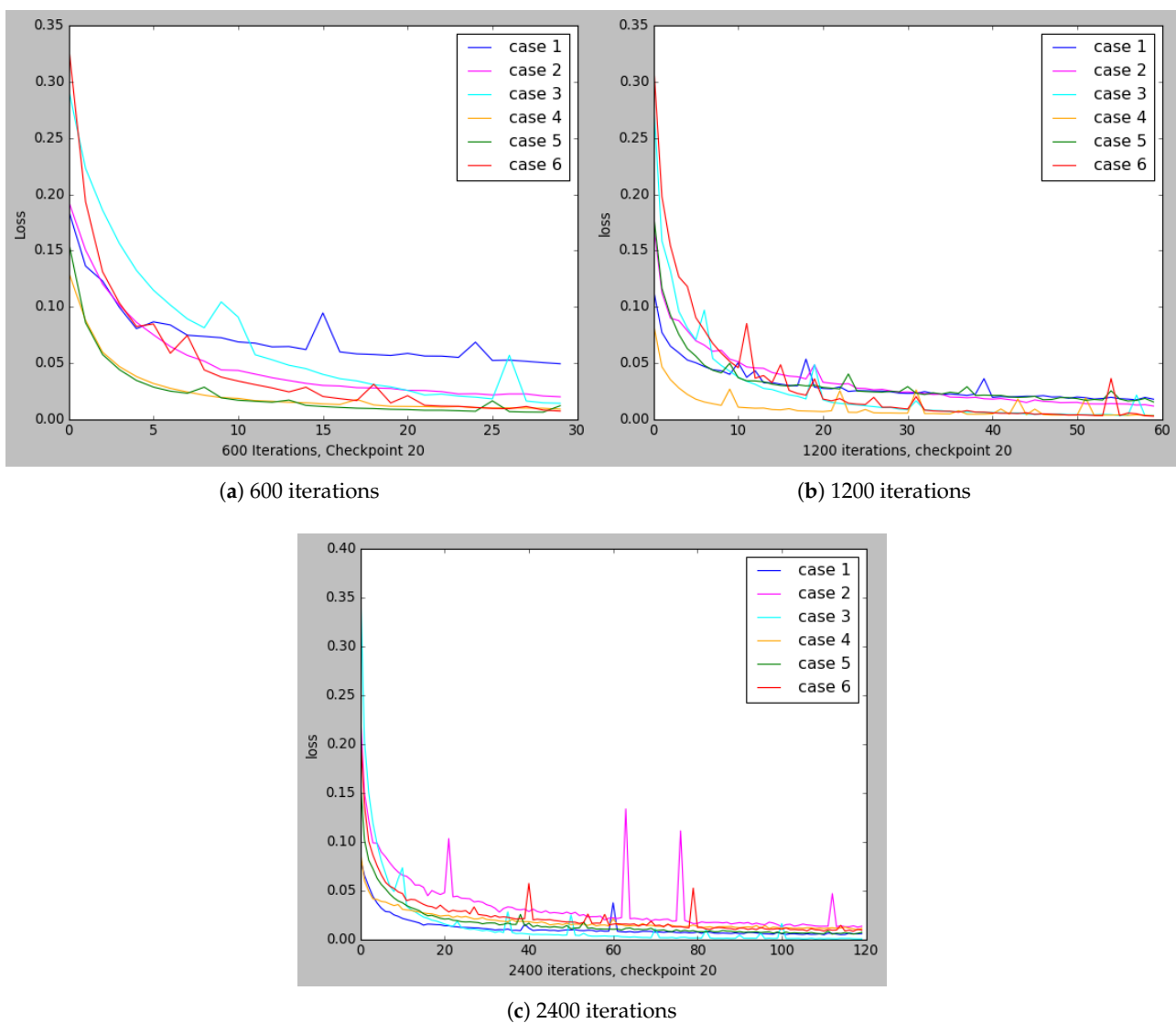
Figure 6. Loss function comparison on cases from Table 5 with different iterations.

(**a**) 600 iterations

(**b**) 1200 iterations

(**c**) 2400 iterations
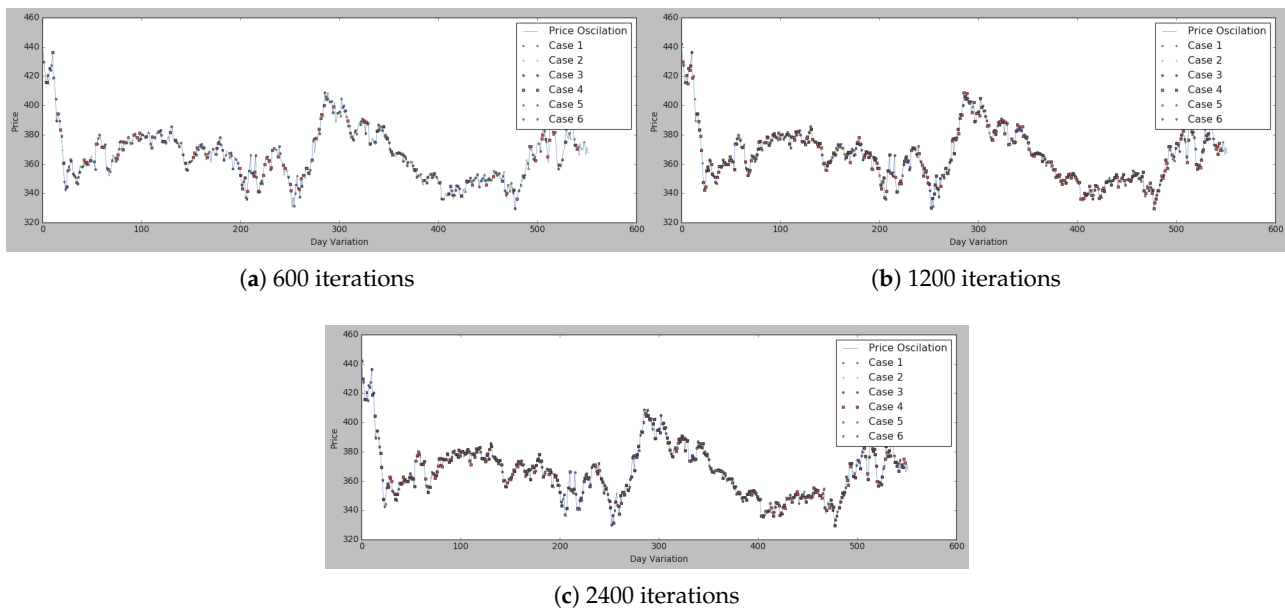
Figure 7. Lowest price identification with different iterations on the basis of the parameters described in Table 5 for each case.

(**a**) Case 1

(**b**) Case 3

(**c**) Case 4
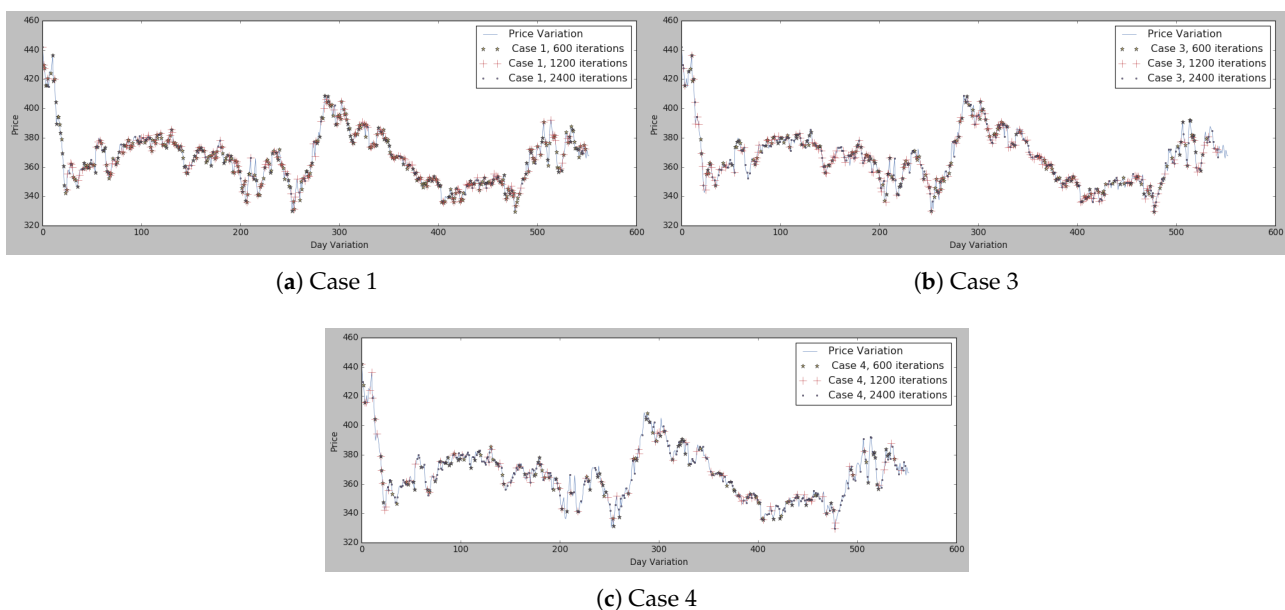
Figure 8. Table 5 for three selected cases

To view the comparisons of cases 1, 3, and 4 in more detail, the results are represented in Figure 7, where all the different cases are compared and represented according to the number of iterations in Figure 7a–c. Has been collected the different cases that were performing better according to the loss functions of Figure 6.

The conclusion that can be drawn from Figures 6 and 7 is that the best options can be chosen by basing the decision-making on hyperparameters, either case 1, 2, or 3, so this case's comparisons have been compared in Figure 8. As can be seen in Figure 8a, case 1 presents good results either in case of 1200 iterations and 2400. Therefore, the option that included in the MAS is the one with the above-mentioned hyperparameters. Thus, the buyer is at last presented with a potential list of suppliers and the lowest market prices at a given moment, which allows them to identify the most suitable suppliers for them as well as the best times to buy the products and also to identify the trends in specific products.

5. Discussion

The use of new technologies as well as the reduction of natural resource consumption or the generation of greenhouse effect impacts is becoming a more important element when choosing a supplier [77]. In the agricultural market, wholesalers purchase and store products in a well-controlled environment in the harvest season, and then recover selected quantities to sell in the market. The amount that is purchased in the harvest season, as well as the amount that is recovered in each selling period, have a strong impact on a wholesaler's profit [78]. The proposed MAS allow to buy agricultural products sustainably thanks to the use of technologies such as edge computing, which reduces agricultural costs and help to make efficient use of resources [16] such as water [20,79] or energy optimization [21], as well as monitoring harmful emissions. According to the work in [78], an optimal selling policy can increase the expected profit. Therefore, being able to combine and obtain real-time price information allows adjusting purchase prices in anticipation of the problem of making strategic sales decisions [80]. In the current case, a paradigm in which each stakeholder (i.e., direct FMCG chains, wholesalers, etc.) acts individually is proposed. For a further evaluation of the model, a pool of suppliers will be contacted in order to include location as another interesting factor when determining prices and creating a reality between the requirements of potential buyers and suppliers.

Author Contributions: M.E.P.-P. and R.S.A. designed and developed the MAS architecture. M.E.P.-P., R.S.A., and O.G. reviewed the state-of-the-art and conducted the case study. G.M. and J.M.C. formalized the problem and reviewed the work. All the authors contributed to the redaction of the paper. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nelson, G.C.; Rosegrant, M.W.; Koo, J.; Robertson, R.; Sulser, T.; Zhu, T.; Ringler, C.; Msangi, S.; Palazzo, A.; Batka, M.; et al. *Climate Change: Impact on Agriculture and Costs of Adaptation*; International Food Policy Research Institute: Washington, DC, USA, 2009; Volume 21.
2. Stevanovic, M.; Popp, A.; Bodirsky, B.L.; Humpenöder, F.; Müller, C.; Weindl, I.; Dietrich, J.P.; Lotze-Campen, H.; Kreidenweis, U.; Rolinski, S.; et al. Mitigation strategies for greenhouse gas emissions from agriculture and land-use change: Consequences for food prices. *Environ. Sci. Technol.* **2017**, *51*, 365–374. [CrossRef]
3. Bellmann, C.; Hepburn, J. The decline of commodity prices and global agricultural trade negotiations: A game changer? *Int. Dev. Policy Rev. Int. Polit. Dev.* **2017**. . [CrossRef]
4. D'Adamo, I.; Falcone, P.M.; Morone, P. A new socio-economic indicator to measure the performance of bioeconomy sectors in Europe. *Ecol. Econ.* **2020**, *176*, 106724. [CrossRef]
5. De Gorter, H.; Drabik, D.; Just, D.R. *The Economics of Biofuel Policies: Impacts on Price Volatility in Grain and Oilseed Markets*; Springer: Berlin/Heidelberg, Germany, 2015.
6. Du, X.; Cindy, L.Y.; Hayes, D.J. Speculation and volatility spillover in the crude oil and agricultural commodity markets: A Bayesian analysis. *Energy Econ.* **2011**, *33*, 497–503. [CrossRef]
7. Kuruppuarachchi, D.; Premachandra, I. Information spillover dynamics of the energy futures market sector: A novel common factor approach. *Energy Econ.* **2016**, *57*, 277–294. [CrossRef]
8. Kang, S.H.; McIver, R.; Yoon, S.M. Dynamic spillover effects among crude oil, precious metal, and agricultural commodity futures markets. *Energy Econ.* **2017**, *62*, 19–32. [CrossRef]
9. Agovino, M.; Casaccia, M.; Ciommi, M.; Ferrara, M.; Marchesano, K. Agriculture, climate change and sustainability: The case of EU-28. *Ecol. Indic.* **2019**, *105*, 525–543. [CrossRef]
10. Eshel, G.; Shepon, A.; Makov, T.; Milo, R. Land, irrigation water, greenhouse gas, and reactive nitrogen burdens of meat, eggs, and dairy production in the United States. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 11996–12001. [CrossRef]
11. Sisinni, E.; Saifullah, A.; Han, S.; Jennehag, U.; Gidlund, M. Industrial Internet of Things: Challenges, Opportunities, and Directions. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4724–4734. [CrossRef]

12. Skobelev, P.; Budaev, D.; Gusev, N.; Voschuk, G. Designing Multi-Agent Swarm of UAV for Precise Agriculture. In Proceedings of the International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS 2018), Toledo, Spain, 20–22 June 2018; pp. 47–59.
13. Zhang, N.; Wang, M.; Wang, N. Precision agriculture—A worldwide overview. *Comput. Electron. Agric.* **2002**, *36*, 113–132. [CrossRef]
14. Sittón-Candanedo, I.; Alonso, R.S.; García, Ó.; Muñoz, L.; Rodríguez-González, S. Edge computing, iot and social computing in smart energy scenarios. *Sensors* **2019**, *19*, 3353. [CrossRef]
15. Alonso, R.S.; Sittón-Candanedo, I.; García, Ó.; Prieto, J.; Rodríguez-González, S. An intelligent Edge-IoT platform for monitoring livestock and crops in a dairy farming scenario. *Ad Hoc Netw.* **2020**, *98*, 102047. [CrossRef]
16. Pérez-Pons, M.E.; Plaza-Hernández, M.; Alonso, R.S.; Parra-Domínguez, J.; Prieto, J. Increasing Profitability and Monitoring Environmental Performance: A Case Study in the Agri-Food Industry through an Edge-IoT Platform. *Sustainability* **2021**, *13*, 283. [CrossRef]
17. Schreinemachers, P.; Berger, T. Land use decisions in developing countries and their representation in multi-agent systems. *J. Land Use Sci.* **2006**, *1*, 29–44. [CrossRef]
18. Gaudou, B.; Sibertin-Blanc, C.; Therond, O.; Amblard, F.; Auda, Y.; Arcangeli, J.P.; Balestrat, M.; Charron-Moirez, M.H.; Gondet, E.; Hong, Y.; et al. The MAELIA multi-agent platform for integrated analysis of interactions between agricultural land-use and low-water management strategies. In Proceedings of the International Workshop on Multi-Agent Systems and Agent-Based Simulation, Saint Paul, MN, USA, 6–7 May 2013; pp. 85–100.
19. García, L.; Parra, L.; Jimenez, J.M.; Lloret, J.; Lorenz, P. IoT-based smart irrigation systems: An overview on the recent trends on sensors and IoT systems for irrigation in precision agriculture. *Sensors* **2020**, *20*, 1042. [CrossRef]
20. González-Briones, A.; Castellanos-Garzón, J.A.; Mezquita-Martín, Y.; Prieto, J.; Corchado, J.M. A multi-agent system framework for autonomous crop irrigation. In Proceedings of the 2nd International Conference on Computer Applications & Information Security (ICCAIS 2019), Riyadh, Saudi Arabia, 1–3 May 2019; pp. 1–6.
21. González-Briones, A.; De La Prieta, F.; Mohamad, M.S.; Omatu, S.; Corchado, J.M. Multi-agent systems applications in energy optimization problems: A state-of-the-art review. *Energies* **2018**, *11*, 1928. [CrossRef]
22. Wang, L.; Feng, J.; Sui, X.; Chu, X.; Mu, W. Agricultural product price forecasting methods: Research advances and trend. *Br. Food J.* **2020**, *122*, 2121–2138. [CrossRef]
23. Valluri, A.; Croson, D.C. Agent learning in supplier selection models. *Decis. Support Syst.* **2005**, *39*, 219–240. [CrossRef]
24. Thies, C.; Kieckhäfer, K.; Spengler, T.S.; Sodhi, M.S. Operations research for sustainability assessment of products: A review. *Eur. J. Oper. Res.* **2019**, *274*, 1–21. [CrossRef]
25. Camargo, G.G.; Ryan, M.R.; Richard, T.L. Energy use and greenhouse gas emissions from crop production using the farm energy analysis tool. *BioScience* **2013**, *63*, 263–273. [CrossRef]
26. Ghadimi, P.; Toosi, F.G.; Heavey, C. A multi-agent systems approach for sustainable supplier selection and order allocation in a partnership supply chain. *Eur. J. Oper. Res.* **2018**, *269*, 286–301. [CrossRef]
27. Azadnia, A.H.; Saman, M.Z.M.; Wong, K.Y. Sustainable supplier selection and order lot-sizing: An integrated multi-objective decision-making process. *Int. J. Prod. Res.* **2015**, *53*, 383–408. [CrossRef]
28. Ghadimi, P.; Wang, C.; Lim, M.K.; Heavey, C. Intelligent sustainable supplier selection using multi-agent technology: Theory and application for Industry 4.0 supply chains. *Comput. Ind. Eng.* **2019**, *127*, 588–600. [CrossRef]
29. Pardoe, D.; Stone, P. Tactex-05: A Champion Supply Chain Management Agent. In Proceedings of the 21st National Conference on Artificial Intelligence (AAAI 2006), Boston, MA, USA, 16 July 2006; pp. 1489–1494.
30. Ganesh, S.; Vadori, N.; Xu, M.; Zheng, H.; Reddy, P.; Veloso, M. Reinforcement learning for market making in a multi-agent dealer market. *arXiv* **2019**, arXiv:1911.05892.
31. Ketter, W.; Collins, J.; Gini, M.; Gupta, A.; Schrater, P. Real-time tactical and strategic sales management for intelligent agents guided by economic regimes. *Inf. Syst. Res.* **2012**, *23*, 1263–1283. [CrossRef]
32. Hogenboom, A.; Ketter, W.; Van Dalen, J.; Kaymak, U.; Collins, J.; Gupta, A. Adaptive Tactical Pricing in Multi-Agent Supply Chain Markets Using Economic Regimes. *Decis. Sci.* **2015**, *46*, 791–818. [CrossRef]
33. Wellman, M.P.; Wurman, P.R. Market-aware agents for a multiagent world. *Robot. Auton. Syst.* **1998**, *24*, 115–125. [CrossRef]
34. Wooldridge, M.J.; Jennings, N.R. Intelligent agents: Theory and practice. *Knowl. Eng. Rev.* **1995**, *10*, 115–152. [CrossRef]
35. Brusaferri, A.; Matteucci, M.; Portolani, P.; Vitali, A. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. *Appl. Energy* **2019**, *250*, 1158–1175. [CrossRef]
36. Maes, P. The agent network architecture (ANA). *ACM Sigart Bull.* **1991**, *2*, 115–120. [CrossRef]
37. Bockstaller, C.; Guichard, L.; Makowski, D.; Aveline, A.; Girardin, P.; Plantureux, S. Agri-environmental indicators to assess cropping and farming systems. A review. *Agron. Sustain. Dev.* **2008**, *28*, 139–149. [CrossRef]
38. Ma, F.; Wei, Y.; Liu, L.; Huang, D. Forecasting realized volatility of oil futures market: A new insight. *J. Forecast.* **2018**, *37*, 419–436. [CrossRef]
39. Tong, Q.; Anders, S.; Zhang, J.; Zhang, L. The roles of pollution concerns and environmental knowledge in making green food choices: Evidence from Chinese consumers. *Food Res. Int.* **2020**, *130*, 108881. [CrossRef]

40. Asemani, M.; Abdollahei, F.; Jabbari, F. Understanding IoT platforms: Towards a comprehensive definition and main characteristic description. In Proceedings of the 5th International Conference on Web Research (ICWR 2019), Tehran, Iran, 24–25 April 2019; pp. 172–177.

41. Mezquita, Y.; González-Briones, A.; Casado-Vara, R.; Chamoso, P.; Prieto, J.; Corchado, J.M. Blockchain-based architecture: A MAS proposal for efficient agri-food supply chains. In Proceedings of the 10th International Symposium on Ambient Intelligence, Ávila, Spain, 26–28 June 2019; pp. 89–96.

42. Wu, C.; Toosi, A.N.; Buyya, R.; Ramamohanarao, K. Hedonic pricing of cloud computing services. *IEEE Trans. Cloud Comput.* **2018**, *9*, 182–196. [CrossRef]

43. Wang, H.; Olhofer, M.; Jin, Y. A mini-review on preference modeling and articulation in multi-objective optimization: Current status and challenges. *Complex Intell. Syst.* **2017**, *3*, 233–245. [CrossRef]

44. Jennings, N.R.; Sycara, K.; Wooldridge, M. A roadmap of agent research and development. *Auton. Agents Multi Agent Syst.* **1998**, *1*, 7–38. [CrossRef]

45. Corchado, J.M.; Laza, R. Constructing deliberative agents with case-based reasoning technology. *Int. J. Intell. Syst.* **2003**, *18*, 1227–1241. [CrossRef]

46. Dignum, V. A Model for Organizational Interaction: Based on Agents, Founded in Logic. SIKS. 2004. Available online: https://dspace.library.uu.nl/bitstream/handle/1874/890/full.pdf?sequence=2 (accessed on 1 May 2021).

47. Esteva, M.; Rodriguez-Aguilar, J.A.; Sierra, C.; Garcia, P.; Arcos, J.L. On the formal specification of electronic institutions. In *Agent Mediated Electronic Commerce*; Springer: Berlin/Heidelberg, Germany, 2001; pp. 126–147.

48. Hübner, J.F.; Sichman, J.S.; Boissier, O. A model for the structural, functional, and deontic specification of organizations in multiagent systems. In Proceedings of the Brazilian Symposium on Artificial Intelligence (SBIA 2002), Porto de Galinhas/Recife, Brazil, 11–14 November 2002; pp. 118–128.

49. Vázquez-Salceda, J. The role of norms and electronic institutions in multi-agent systems applied to complex domains. The HARMONIA framework. *Ai Commun.* **2003**, *16*, 209–212.

50. Zambonelli, F.; Jennings, N.R.; Wooldridge, M. Developing multiagent systems: The Gaia methodology. *ACM Trans. Softw. Eng. Methodol. TOSEM* **2003**, *12*, 317–370. [CrossRef]

51. Kuhn, H.W.; Tucker, A.W. *Contributions to the Theory of Games*; Princeton University Press: Princeton, NJ, USA, 1953; Volume 2.

52. Binmore, K. Foundations of game theory. In *Advances in Economic Theory: Sixth World Congress*; Cambridge University Press: Cambridge, UK, 1992; Volume 1, pp. 1–31.

53. Smith, J.M. Evolutionary game theory. In *Vito Volterra Symposium on Mathematical Models in Biology*; Springer: Berlin/Heidelberg, Germany, 1980; pp. 73–81.

54. Smith, J.M.; Price, G.R. The logic of animal conflict. *Nature* **1973**, *246*, 15–18. [CrossRef]

55. Parkes, D.C.; Wellman, M.P. Economic reasoning and artificial intelligence. *Science* **2015**, *349*, 267–272. [CrossRef]

56. Marreiros, G.; Novais, P.; Machado, J.; Ramos, C.; Neves, J. An agent-based approach to group decision simulation using argumentation. In Proceedings of the International MultiConference on Computer Science and Information Tecnology, Workshop Agent-Based Computing (ABC 2006), Wisla, Poland, 6–10 October 2006; pp. 225–232.

57. Sadeghi, A.; Zandieh, M. A game theory-based model for product portfolio management in a competitive market. *Expert Syst. Appl.* **2011**, *38*, 7919–7923. [CrossRef]

58. Ban, G.Y.; El Karoui, N.; Lim, A.E. Machine learning and portfolio optimization. *Manag. Sci.* **2018**, *64*, 1136–1154. [CrossRef]

59. Balakrishnan, P.; Jacob, V.S. Genetic algorithms for product design. *Manag. Sci.* **1996**, *42*, 1105–1117. [CrossRef]

60. Freitas, F.D.; De Souza, A.F.; de Almeida, A.R. Prediction-based portfolio optimization model using neural networks. *Neurocomputing* **2009**, *72*, 2155–2170. [CrossRef]

61. Ma, Y.; Han, R.; Wang, W. Prediction-Based Portfolio Optimization Models Using Deep Neural Networks. *IEEE Access* **2020**, *8*, 115393–115405. [CrossRef]

62. Almahdi, S.; Yang, S.Y. An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. *Expert Syst. Appl.* **2017**, *87*, 267–279. [CrossRef]

63. Van Calker, K.; Berentsen, P.; Romero, C.; Giesen, G.; Huirne, R. Development and application of a multi-attribute sustainability function for Dutch dairy farming systems. *Ecol. Econ.* **2006**, *57*, 640–658. [CrossRef]

64. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.

65. Watkins, C.J.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]

66. Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep reinforcement learning: A brief survey. *IEEE Signal Process. Mag.* **2017**, *34*, 26–38. [CrossRef]

67. Watkins, C.J.C.H. Learning from Delayed Rewards. Ph.D. Thesis, University of Cambridge, UK, 1989.

68. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing atari with deep reinforcement learning. *arXiv* **2013**, arXiv:1312.5602.

69. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef]

70. Bradtke, S.J.; Duff, M.O. Reinforcement learning methods for continuous-time Markov decision problems. *Adv. Neural Inf. Process. Syst.* **1995**, *7*, 393–400.

71. Sutton, R.S. Learning to predict by the methods of temporal differences. *Mach. Learn.* **1988**, *3*, 9–44. [CrossRef]

72. Hasselt, H. Double Q-learning. *Adv. Neural Inf. Process. Syst.* **2010**, *23*, 2613–2621.
73. Oliveira, E.; Fischer, K.; Stepankova, O. Multi-agent systems: Which research for which applications. *Robot. Auton. Syst.* **1999**, *27*, 91–106. [CrossRef]
74. Gregori, M.E.; Cámara, J.P.; Bada, G.A. A jabber-based multi-agent system platform. In Proceedings of the 5th International Joint Conference on Autonomous Agents and Multiagent Systems, Hakodate, Japan, 8–12 May 2006; pp. 1282–1284.
75. Atsalakis, G.S.; Valavanis, K.P. Surveying stock market forecasting techniques–Part II: Soft computing methods. *Expert Syst. Appl.* **2009**, *36*, 5932–5941. [CrossRef]
76. Schmidt-Hieber, J. Nonparametric regression using deep neural networks with ReLU activation function. *Ann. Stat.* **2020**, *48*, 1875–1897.
77. Gurel, O.; Acar, A.Z.; Onden, I.; Gumus, I. Determinants of the green supplier selection. *Procedia Soc. Behav. Sci.* **2015**, *181*, 131–139. [CrossRef]
78. Liu, H.; Zhang, J.; Zhou, C.; Ru, Y. Optimal purchase and inventory retrieval policies for perishable seasonal agricultural products. *Omega* **2018**, *79*, 133–145. [CrossRef]
79. Villarrubia, G.; Paz, J.F.D.; Iglesia, D.H.; Bajo, J. Combining multi-agent systems and wireless sensor networks for monitoring crop irrigation. *Sensors* **2017**, *17*, 1775. [CrossRef]
80. Shi, J.; Zhao, Y.; Kiwanuka, R.B.K.; Chang, J. Optimal Selling Policies for Farmer Cooperatives. *Prod. Oper. Manag.* **2019**, *28*, 3060–3080. [CrossRef]