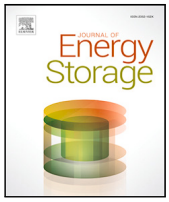
Journal of Energy Storage 93 (2024) 112031

Contents lists available at ScienceDirect

Journal of Energy Storage

journal homepage: www.elsevier.com/locate/est



Research papers

A reinforcement learning approach to dairy farm battery management using Q learning

Nawazish Ali [a,*], Abdul Wahid [a,b], Rachael Shaw [c], Karl Mason [a]

[a] School of Computer Science, University of Galway, Galway, H91 FYH2, Ireland
[b] Department of Computer Science and Engineering, Indian Institute of Information Technology, Dharwad, 580009, India
[c] Atlantic Technological University, Galway, H91 T8NW, Ireland

ARTICLE INFO

Keywords:
Reinforcement learning
Dairy farming
Battery management
Q-learning

ABSTRACT

Dairy farming consumes a significant amount of energy, making it an energy-intensive sector within agriculture. Integrating renewable energy generation into dairy farming could help address this challenge. However, fluctuations in renewable generation pose a challenge to this integration. Effective battery management techniques are needed to store and utilize the energy generated from renewable sources. The objective of this research is to leverage Reinforcement Learning to develop an effective approach for battery management systems in dairy farming. Our work contributes by implementing a Q-learning algorithm for dairy farm battery management, incorporating wind and solar energy, exploring the state space of the algorithm, and considering Ireland as a case study as it works towards attaining its 2030 energy strategy centered on the utilization of renewable sources. The findings show that the proposed algorithm reduces the cost of imported electricity from the grid by 13.41%, 24.49% when utilizing wind generation, and peak demand by 2%. These findings highlight the effectiveness of Reinforcement Learning for battery management in the dairy farming sector.

1. Introduction

The growth in the global population has led to an increased demand for food products. Milk maintains an important place in the dietary patterns of individuals across the globe because of its essential nutritious components. According to the Food and Agriculture Organisation (FAO), global milk production rose from 735 million metric tonnes in 2000 to 855 million metric tonnes in 2019 [1]. The rising demand for dairy products has led to an increase in the number of dairy farms [2]. These farms heavily rely on electricity for multiple activities like milk cooling, water heating, pumping, and lighting [3]. Meeting these energy needs requires substantial imports of electricity from external grids. However, the rising cost of electricity necessitates considering alternative energy sources like solar photovoltaic and wind turbines. Embracing renewable energy sources can help satisfy the energy requirements of farms and decrease dependence on the external grid for importing electricity [4]. By 2030 Irish government aims to transition to a low-carbon economy within the EU, emphasizing renewable energy, secure electricity supply, and enhanced energy efficiency [5]. This research supports these goals by efficiently managing energy storage, increasing renewable energy use, and reducing carbon emissions by minimizing grid reliance as the grid generates most of the energy by burning fossil fuels [6].

Power generated from renewable energy sources exhibits temporal variability. Employing a battery management system for storing electrical energy is crucial for future use. Different battery management systems are applied to different applications [7]. The use of batteries has the potential to influence the economic aspects of electricity consumption within dairy farming. However, optimizing battery performance necessitates the implementation of different strategies. Various conventional methods have been employed to improve battery efficiency. These approaches include such as Maximizing Self-Consumption (MSC) and Time of Use (TOU) [8].

In recent years, there has been remarkable progress in Artificial Intelligence (AI), largely driven by the data revolution. This advancement has shown immense potential across various fields, yielding promising results [9]. One significant area within AI is Reinforcement Learning (RL), where agents can operate in stochastic environments without explicit knowledge of the environment or predefined decision-making processes. Instead, with established objective functions. Two main algorithms that are particularly prominent in the field of RL are actor-critic learning and Q-learning [10,11]. RL agents can effectively learn a policy in diverse domains through these algorithms. This research aims to use RL agents to learn battery management policies efficiently. The primary goal is to optimize the charging/discharging of a battery

* Corresponding author.
E-mail address: N.Ali3@universityofgalway.ie (N. Ali).

https://doi.org/10.1016/j.est.2024.112031
Received 30 January 2024; Received in revised form 19 April 2024; Accepted 12 May 2024
Available online 30 May 2024

to maximize renewable utilization and reduce the cost of electricity imported from the grid.

Existing research has highlighted the efficacy of RL in battery management across diverse contexts. However, the integration of RL into the agricultural sector, specifically in dairy farming for battery management, remains relatively unexplored in current literature. The main contributions of this research are specified below:

1. Using Q-learning we present an autonomous learning approach for efficient battery management in dairy farms, and we demonstrate the effectiveness of our algorithm in achieving improved energy efficiency over established baseline algorithms.
2. In contrast to existing approaches, this work also analyses the influence of incorporating wind energy data on the effectiveness of battery management.
3. We evaluate the impact of variations in state space information on the performance of our Q-learning approach. We explore the impact of additional parameters including load, PV, and wind to determine the most optimal solution.
4. We extend our experimental analysis to also evaluate the Q-Learning algorithm's performance using data based on case studies in Finland and Ireland, focusing on dairy farm battery management.

The remainder of this paper is structured as follows: Section 2 examines both conventional and AI methodologies in battery management. Section 3 formulates the research problem and the proposed methodology. Section 4 Evaluate the performance of our proposed approach. Finally, Section 5 concludes the research, emphasizing the primary contributions of this work.

2. Background and related work

Numerous researchers have worked on improving battery management to reduce reliance on external power grids and lower electricity import expenses. Surprisingly, the use of RL in battery management within the dairy farming sector has been unexplored. This study employs RL techniques to manage batteries in dairy farming, to reduce dependence on external power grids. Researchers investigated different methods for efficiently handling battery management, including conventional battery control methods such as rule-based and dynamic programming strategies, as well as AI methods, mainly RL. There is a rising interest in utilizing RL, particularly because of the volatile nature of the environment within agent interact. RL agents can adapt to volatile or non-deterministic environments.

2.1. Reinforcement learning

RL is utilized in the research to effectively manage the battery controller to maximize the utilization of renewable generation. RL involves interaction between an agent and its environment to maximize the cumulative reward obtained from the environment through specific actions taken by the agent. The environment can be characterized as a Markov Decision Process (MDP). An MDP comprises a state space denoted by $\boldsymbol{S}$, an action space denoted by $\boldsymbol{A}$, a state change denoted by $p(s_t + 1|s_t; a_t)$ where p represents a probability distribution governing state transitions, and a reward function denoted by $R : S \times A \rightarrow R$. The agent takes actions at each time interval based on the current state observations. The agent changes behavior by considering the outcomes and feedback from previous actions. A policy determines how the agent acts in the environment denoted as π. The function π maps each state in a given environment to a probability distribution of possible actions. The reward from a state is defined as the sum of discounted future rewards, which can be mathematically represented as $R_t = \sum_{i=t}^{T} \gamma^{(i-t)} r(S_i, A_i)$. In this equation, R_t defines the reward at time $\boldsymbol{t}$. The symbol γ represents the discount factor, ranging from 0 to 1, and $\pi - t$ illustrates the importance of future rewards as compared to immediate rewards. The rewards are influenced by the actions taken, which are determined by the policy π. The goal of RL is to develop a policy that maximizes the expected cumulative reward starting from the initial probability distribution. The aim is to maximize the total reward received from the environment.

The expected result of acting in a specific state, while obtaining a certain policy, is calculated using the action-value function. Eq. (1), a fundamental component in many RL algorithms, provides a means to evaluate the potential outcome of actions within the framework of the given policy.

$$Q^{\pi}(s_t, a_t) = E_{\pi}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t, a_t\right] \tag{1}$$

Eq. (1) denotes the action-value function for a given policy π at a specific time t, concerning the state s_t and action a_t. The returned value represents the expected cumulative discounted reward for taking action at in state s_t and then following policy π for all future time steps. The symbol E_{π} denotes the expected value under policy π. Additionally, the summation $\sum_{k=0}^{\infty}$ denotes the sum over all potential future time steps that begin from time $t+1$. The variable r_{t+k+1} denotes the reward acquired at time t+k+1 subsequent to executing the action a_t within the state s_t. The γ represents the discount factor which maximizes the future reward.

Q-learning is one of the basic RL algorithms that does not require a model of the environment. It is commonly used to determine the best policy for selecting actions in a finite Markov decision process. This approach aims to get information regarding the significance of an action within a specific state, enabling an agent to make decisions that optimize the overall accumulated reward over a given period. The algorithm comprises the process of updating Q-values, which are action-value pairs, that are stored within a table. Each Q-value corresponds to the anticipated utility of executing a specific action within a particular condition, subsequently following the optimal policy. The main formulation to update the Q-value is depicted in Eq. (2)

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a} Q(s_{t+1}, a) - Q(s_t, a_t)] \tag{2}$$

The update rule for the Q-value of the current state–action pair (s_t, a_t) involves a scalar factor α, Also known as a learning rate which controls the rate at which the agents will explore the environment ranging from 0 to 1 and scales the difference between the observed reward $\boldsymbol{R}_t$ plus the discounted estimate of the maximum Q-value for the next state s_{t+1} (discounted by a factor γ) and the current estimate of the Q-value for the current state–action pair $Q(s_t, a_t)$.

2.2. Conventional battery control methods

Many studies have focused on identifying effective operational strategies for PV battery systems, for different objectives [12]. Specifically, the Maximizing Self-Consumption (MSC) and Time of Use (TOU) methods for battery charging [13].

The MSC is a method used for managing battery charging and discharging by maximizing the utilization of solar power generation. It charges the maximum amount of solar energy available [12]. Braun et al. highlighted that optimal battery usage significantly increases the local consumption of solar energy [14]. Further, they investigated the optimal sizing of photovoltaic systems [15] and the capacity requirements for energy storage [16], aiming to maximize the use of locally generated solar power and reduce reliance on external power grids. In their comprehensive review, Luthander et al. analyzed previous studies on solar power self-consumption in buildings, concluding that proper battery sizing can improve self-consumption rates by 13%–24% [17]. Sharma et al. conducted a study on the optimization of battery size for zero-net energy homes equipped with rooftop solar panels in South Australia, employing the MSC operational strategy [18]. Their findings suggest that installing suitable batteries can enhance the self-consumption of solar energy by 20%–50% [19].

TOU strategy uses electricity prices for charging and discharging the battery; it charges the battery when the prices are low and discharges at peak times. Feed-in Tariffs (FiT) and TOU pricing strategies, implemented in several countries, aim to enhance the adoption of Photovoltaic Battery (PVB) systems and encourage consumer involvement in energy management, which has been a significant area of research. [20]. Other studies have focused on the TOU tariff method for efficient battery management. For instance, Christoph et al. utilized optimization techniques to refine the TOU rate structure [21], while Li et al. developed TOU tariffs using the Gaussian Mixture Model [22]. This approach has enabled prosumers to get economic advantages by taking advantage of FiT and adapting to varying electricity prices during peak and off-peak times, which is a main benefit of the TOU strategy [23]. Research by Gitizadeh et al. and Hassan et al. explores optimizing battery capacity, by utilizing TOU [24,25]. Additionally, Ratnam et al. found that many PVB system users were able to achieve significant annual cost reductions through FiT programs [26].

2.3. Reinforcement learning for energy management

RL algorithms are widely used in various applications. Wei et al. implemented dual iterative Q-learning for managing batteries in smart residential settings [27]. This form of Q-learning is designed to enhance energy management in smart homes by optimizing the charging and discharging of batteries. Similarly, Kim et al. developed an RL-based algorithm for energy management in smart buildings [28]. Their approach uses RL to dynamically identify the most effective energy regulation strategy based on real-time data. Ruelens et al. also applied RL, but to the operation of an electric water heater [29], using the algorithm to boost the heater's energy efficiency by learning and adapting to real-time user demand and grid conditions. Research indicates that RL can significantly enhance both efficiency and cost-effectiveness in energy consumption within smart grids. Furthermore, Li et al. introduced a multi-grid RL method to optimize the energy efficiency and comfort of Heating, Ventilation, and Air Conditioning (HVAC) systems [30]. This method balances HVAC energy use with maintaining optimal room temperature and humidity. Their findings suggest that this approach effectively optimizes energy consumption while ensuring comfortable indoor environments.

2.4. Reinforcement learning for battery management

Numerous studies have explored battery management using RL. Foruzan et al. introduced the use of RL for managing energy in microgrids [31]. They employed an RL system capable of adapting in real-time to changing energy needs and generating renewable energy, enhancing the energy efficiency of microgrids. RL is effective in improving energy consumption in a cost-efficient manner. In a similar application, Guan et al. developed an RL-based solution for controlling domestic energy storage to reduce electricity cost [32]. This RL method optimizes the charging and discharging of energy storage systems, helping decrease peak power demands and shift energy usage to cheaper, off-peak times, lowering electricity bills. Their simulations demonstrated that this strategy could effectively reduce the electricity cost associated with household energy storage systems. Liu et al. explored the use of Deep Reinforcement Learning (DRL) for optimizing energy management in households [33]. This study utilized a DRL system designed to enhance energy efficiency in smart homes by constantly learning the most effective energy management strategies. In simulated smart home environments, this DRL-based approach was more efficient and cost-effective than traditional rule-based methods, indicating its potential to significantly improve energy management in intelligent residential settings. Cao et al. proposed the DRL method for battery charging and discharging, handling power price uncertainty, improving the accuracy of the degradation model, and non-linear charging and discharging efficiency [34]. They demonstrated the algorithm's efficacy and performance by testing it on historical wholesale electricity data from the United Kingdom. Yu et al. use Deep Deterministic Policy Gradient (DDPG) for the home energy management system to minimize electricity cost by scheduling HVAC systems and Effective Solutions for Storing (ESSs) [35]. By leveraging the dynamic prices, the results demonstrated that the proposed algorithm saves energy cost by 8.1%–15.21%.

Abedi et al. have created a real-time intelligent battery energy control system for residential buildings that incorporates solar panels, battery energy systems, and grid connectivity by using Q-learning [36]. The results of their study demonstrate that the algorithm effectively decreases the monthly electricity cost by 7.99% to 3.63% for house 27 and 6.91% to 3.26% for house 387. Wei et al. proposed the DDPG a DRL algorithm for the fast charging of lithium-ion batteries (LIB) [37]. They compare the proposed algorithm with the rule base by considering different constraints i.e. LIB temperature, charging rapidity, and degradation. Huang et al. introduce Proximal Policy Optimisation (PPO) as a DRL algorithm to optimize the capacity scheduling of solar battery systems [38]. To enhance the safety of the battery, a safety control algorithm is implemented by utilizing a serial approach incorporated with a PPO algorithm. Their findings indicate that the proposed algorithm outperforms other DRL algorithms. Cheng et al. propose a periodic deterministic policy gradient (PDPG) to schedule the charging of multi-battery energy storage systems (MBESS) [39]. Their research shows that compared to the DPG algorithm, the PDPG algorithm reduces power cost by 8.79%. Paudel et al. employ the MDP framework to efficiently manage battery storage systems' charging and discharging operations by considering the electricity price fluctuations and other relevant parameters [40]. The authors substantiate their method's effectiveness by installing 150 fast charging stations and a battery storage system throughout the Pennsylvania-New Jersey-Maryland region. The studies mentioned above show RL's impact on battery management applications.

The conventional and RL studies underscore the importance of maximizing local energy utilization and optimizing battery usage. However, some limitations have been identified in these works that our research aims to address. Firstly, they did not address performance variations under diverse weather conditions and geographical locations, besides the impact of fluctuating energy prices and renewable generation. Secondly, these studies focus solely on one renewable source and do not consider the effects of integrating other energy sources. Lastly, all the conventional and RL methods have been applied in smart homes and buildings, but their adaptation to dairy farm battery management remains largely unexplored. Dairy farms typically consume more energy than households or offices due to operational needs and reliance on high-energy equipment like milking machines and milk cooling systems, which account for 20%–30% of the farm's electricity. Furthermore, research has shown that electricity consumption per dairy cow ranges from 4 to 7.3 kWh/week [3]. In contrast, households and offices use energy mainly for heating, cooling, lighting, and appliances. The unique requirements of dairy farming operations lead to higher load consumption and diverse consumption patterns. This research addresses these gaps by demonstrating how a Q-learning algorithm optimizes battery management in dairy farming settings. It also mitigates drawbacks by testing the proposed methodology across various locations and weather conditions, integrating multiple renewable sources, and considering electricity price fluctuations.

3. Methodology

3.1. System design

The PVB system connected to the grid, as depicted in Fig. 1, includes a set of components: solar panels, a battery storage unit, the power grid, and a dairy farm that utilizes electricity from both solar and

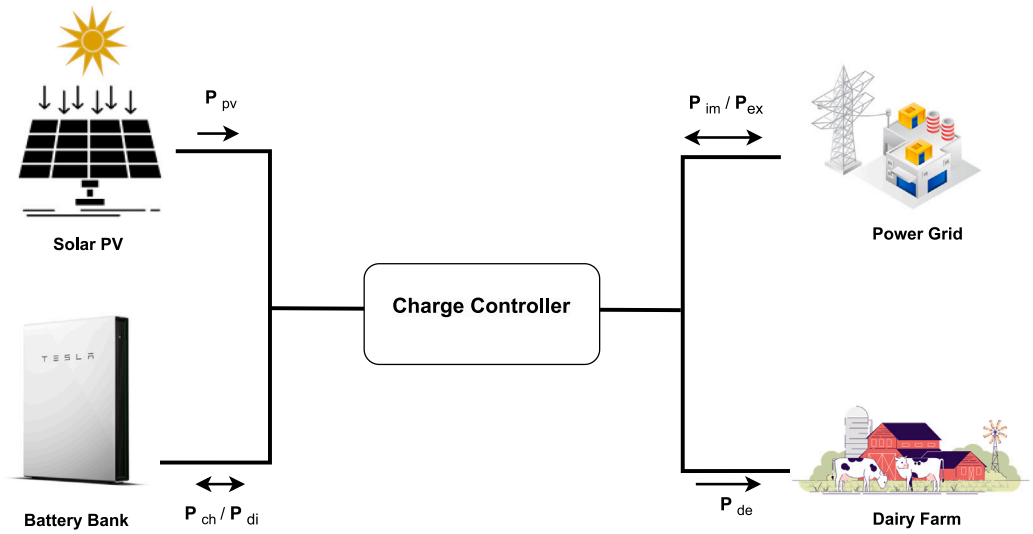


Fig. 1. Overview of the system architecture.

grid sources. The energy storage system considered for this research is the Tesla Powerwall 2.0, which offers a substantial capacity of 13.5 kWh and supports both charging and discharging 5 kW [41]. The PV-generated electricity is used to meet the farm's load, charge the battery, or sell it back to the grid, according to the operational requirements. The role of the charge/discharge controller is to charge and discharge the battery according to the renewable generation, electricity demand, and price of electricity. Meanwhile, the power grid is connected to the dairy farm and the battery. It supplies electricity when there is high demand and low renewable generation. The battery storage is used to satisfy the farm's extra energy needs, a process commonly referred to as peak shaving [42]. This involves using excess energy demands by utilizing stored power in the battery.

3.2. Data and price profile

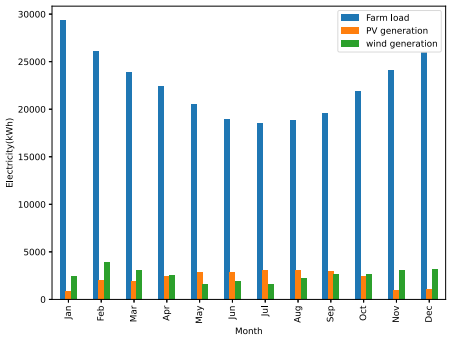
For this study, two datasets were used: One dataset from Finland to train the algorithm and a second dataset collected from Ireland to evaluate the performance of the algorithm. The Finland dataset has information about the load demand from dairy farms, PV generation, wind generation, and electricity prices. The load data is collected from [43] and provides hourly electricity consumption over a year. Fig. 2(a) demonstrates the monthly distribution of electricity demands for a dairy farm and PV generation and wind energy generated by the dairy farm throughout the year. The dataset consists of a dairy farm that has approximately 180 cows and has an estimated annual electricity usage of around 261 megawatt-hours (MWh). The PV and wind data was collected from the System Advisor Model (SAM) having a capacity of 20 kW [44]. The Finland electricity price data was collected from a Helsinki electricity supply company [45]. This price data is dynamic and includes three different price levels [46]. The lowest rate is during off-peak hours, the standard rate applies for most of the day, and a higher peak rate is charged during the busiest hours. Specifically, the pricing is segmented into three time periods. The off-peak hours, with the lowest rate, are from 11 p.m. to 7 a.m. The standard rate applies during two intervals: from 8 a.m. to 5 p.m. and from 7 p.m. to 10 p.m. The peak rate, which is the highest, is charged between 5 p.m. and 7 p.m. Fig. 2(b) shows how these electricity prices fluctuate over the day.

The Ireland dataset includes data on farm load, PV generation, and electricity pricing. However, it lacks wind generation data as we were unable to find wind generation data for Ireland. The load consumption data, detailing electricity from the dairy farm over a year, was collected from a study on Irish dairy farms [47]. The PV generation data was collected from SAM [44] having a capacity of 20 kW. The price data is collected from the Ireland electricity supply company Electric Ireland [48]. Fig. 3 shows the Irish dairy farm energy consumption and photovoltaic (PV) energy generation and electricity price. This figure illustrates the variations in PV generation and electricity price, it also demonstrates the farm electricity demand patterns. The aim is to explore the relationship between energy consumption and PV generation, particularly in the Irish dairy farm context. Fig. 3(a) specifically illustrates the monthly load demand and PV generation of the dairy farm over one year, while Fig. 3(b) illustrates the price variations over the day.

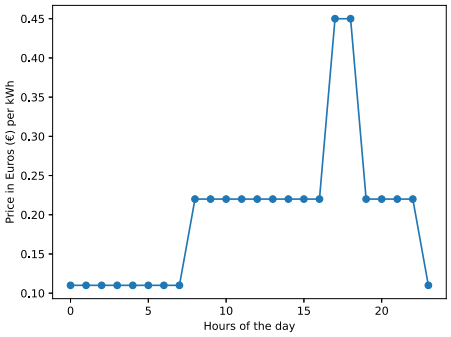
3.3. Baseline battery controllers

The battery management system was optimized through the implementation of two rule-based strategies which are MSC and TOU in the baseline algorithm [8]. The MSC is a means of optimizing the utilization of surplus energy generated by the PV system through its storage in a battery. The TOU involved modifying the battery charging process in response to variations in electricity prices. These two strategies were implemented to manage the battery as a baseline comparison method.

The MSC strategy is a prevalent energy management approach utilized in PV-integrated energy systems. Its primary objective is to optimize the utilization of PV-generated power for load demand and battery charging. The core principle of this system is that when the energy produced by PV sources surpasses the current energy needs, any excess energy is stored in the battery, and the remaining energy is transmitted to the grid. In cases where the PV generation falls short of the required load, the battery is utilized as the primary source for meeting the load demand. It is discharged to ensure that the load demand is met. If the load demand exceeds the combined capacity of the PV system and battery, external electricity will be purchased from the power grid to compensate for the shortfall. The MSC mostly depends on PV generation and if the PV is not available this strategy does not work well. The pseudocode of the MSC strategy is presented in the Algorithm 1.

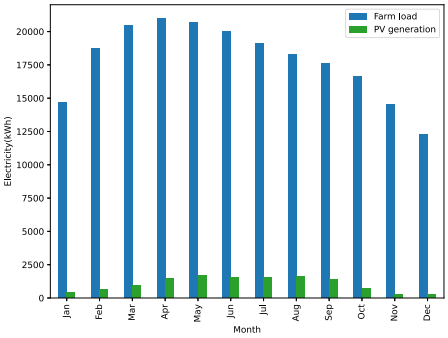


(a) Finland farm load and PV and wind generation for one year.

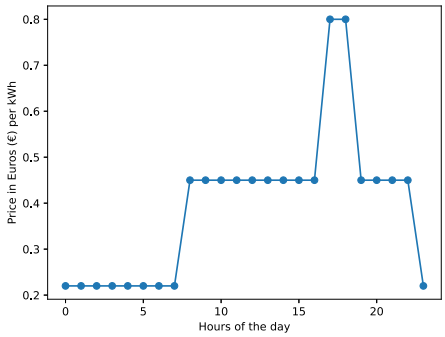


(b) Finland electricity price profile

Fig. 2. Dairy farm electricity, solar photovoltaic generation, and price data from Finland.



(a) Ireland farm load and PV generation for one year.



(b) Ireland electricity price profile

Fig. 3. Dairy farm electricity, solar photovoltaic generation, and price data from Ireland.

Algorithm 1 MSC Strategy for Battery Management

Initialize *pv_generation*, *load_demand*, *battery_capacity*, *total_episodes*
while *episode* = 1 **to** *total_episodes* **do**
  Update *pv_generation* and *load_demand*
  if *pv_generation* > *load_demand* **then**
    excess_energy ← *pv_generation* − *load_demand*
    if *battery_capacity* can store *excess_energy* **then**
      Store *excess_energy* in battery
    else
      Store in battery up to *battery_capacity*
      Transmit remaining *excess_energy* to grid
    end if
  else if *pv_generation* < *load_demand* **then**
    Use battery to meet *load_demand*
  end if
end while

Algorithm 2 TOU Strategy for Battery Management

Define *peak_hours*, *off_peak_hours*, *total_episodes*
Initialize *pv_generation*, *load_demand*, *battery_capacity*, *total_episodes*, *electricity_prices*
while *episode* = 1 **to** *total_episodes* **do**
  Update *pv_generation*, *load_demand*, *current_time*
  if In off_peak_hours and battery not full **then**
    Charge battery from grid at max rate
  end if
  if *pv_generation* > *load_demand* **then**
    Store excess PV in battery
  end if
  if In *peak_hours* and *load_demand* > *pv_generation* **then**
    Use battery to meet shortfall
  end if
end while

The adoption of the TOU strategy is aimed at achieving economic gains through the utilization of the price variation between peak and off-peak electricity rates. The primary objective of the TOU strategy charge the battery during the valley price period and subsequently discharge the stored electricity to meet load demand during high/peak periods. In addition, the TOU strategy charges the battery at the highest possible rate from the grid during the off-peak period (23:00-7:00 the following day). In instances of peak prices, the battery is discharged to fulfill the energy demand of the farm when load demand exceeds the capacity of the photovoltaic generation. The pseudocode of the TOU strategy is presented in the Algorithm 2.

3.4. Q learning

This paper utilizes the Q-learning approach which is an effective RL technique for efficient battery management used by various researchers as described in the literature. The Q-learning algorithm operates by choosing the action that corresponds to the maximum Q-value in each state. Eq. (3) illustrates the maximum Q-value selection strategy.

$$Q^*(s_t, a_t) = argmax_{a \in A} Q^{\pi}(s_t, a_t) \tag{3}$$

The symbol $Q^*(s_t, a_t)$ denotes the optimal action that maximizes the action-value function $Q^{\pi}(s_t, a_t)$ at time t, with respect to the state s_t.

The mathematical symbol $argmax_{a\in A}$ denotes the maximum value of the action-value function across the set of all possible actions belonging to the action space A, given the state s_t. The aforementioned statement implies that the optimal value of the action, denoted by $Q^*(s_t, a_t)$, results in the maximum reward for the agent in the state s_t.

Q-learning algorithms employ the Bellman equation [49] to choose maximum Q-values and the generalized Bellman equation is expressed in Eq. (4).

$$Q_\pi(s,a) = \sum_{s',r} p(s',r|s,a)[r + \gamma \sum_{a'} \pi(a'|s')Q_\pi(s',a')] \tag{4}$$

Eq. (4) presents the correlation between the action-value function Q_π, the reward, and transition probabilities of the environment. The statement specifies that the value of $Q_\pi(s,a)$ is equivalent to the summation of the probability $p(s',r|s,a)$ of transitioning to state $\boldsymbol{S'}$ and receiving reward $\boldsymbol{R}$, multiplied by the summation of the immediate reward $\boldsymbol{R}$ and the discounted value of the subsequent state $\boldsymbol{S'}$ under the policy π, considering all feasible next states $\boldsymbol{S'}$ and rewards $\boldsymbol{R}$. The parameter γ, commonly referred to as the discount factor, plays a crucial role in determining the relative significance of rewards that are obtained immediately versus those that are obtained in the future. The Bellman equation is a fundamental concept within the field of RL, used for numerous algorithms that aim to acquire knowledge regarding the value function and policy optimization.

Q-learning involves using the current estimate of Q^π to improve its future predictions by including the known reward value $r(s_t, a_t)$. Q-learning fundamentally relies on the concept of Temporal Difference (TD) learning [50]. In this method, the Q-value is updated after performing an action in the state S_t and observing the resulting reward r_t which leads to a transition to the next state s_{t+1}. The TD is mathematically represented in Eq. (2).

Empirical evidence supports the notion that as the frequency of visits to each state–action pair's Q-value approaches infinity, the learning rate α exhibits a decreasing trend concerning the time step $\boldsymbol{t}$. As the value of $\boldsymbol{t}$ approaches infinity, the function $Q(s;a)$ approaches the optimal $Q*(s;a)$ for all possible state–action pairs [11]. In this study, the Q-learning algorithm was utilized to optimize the management of battery charging and discharging operations to reduce the cost of imported electricity from the power grid. The Q-learning algorithm comprises different components, namely the state space denoted as S, the action space represented by A, and the reward function, which is the aggregate cost of electricity denoted as R.

3.5. Application of Q-learning to battery management

In this study, Q-learning is employed as a means of effectively managing the process of battery charging and discharging. This is achieved through the exploration of the state space and action space, which are integral components of the environment. The reward is calculated by considering various actions, such as charging, discharging, or remaining idle, in response to factors such as renewable generation and electricity prices. The state space, action space, and reward are explained below. The proposed algorithm is illustrated in the flow chart shown in Fig. 4. The algorithm begins by initializing the environment, specifying the available actions, defining a strategy for computing rewards, and determining the number of episodes. Subsequently, it initializes the learning rate and exploration rate to 0.8, the discount factor to 0.9, and initializes the Q Table to 0. The algorithm employs the weight decay with the decay of 0.0001 to gradually decrease the learning rate and exploration rate concerning the episodes. To determine the appropriate action, the algorithm uses the epsilon-greedy policy.

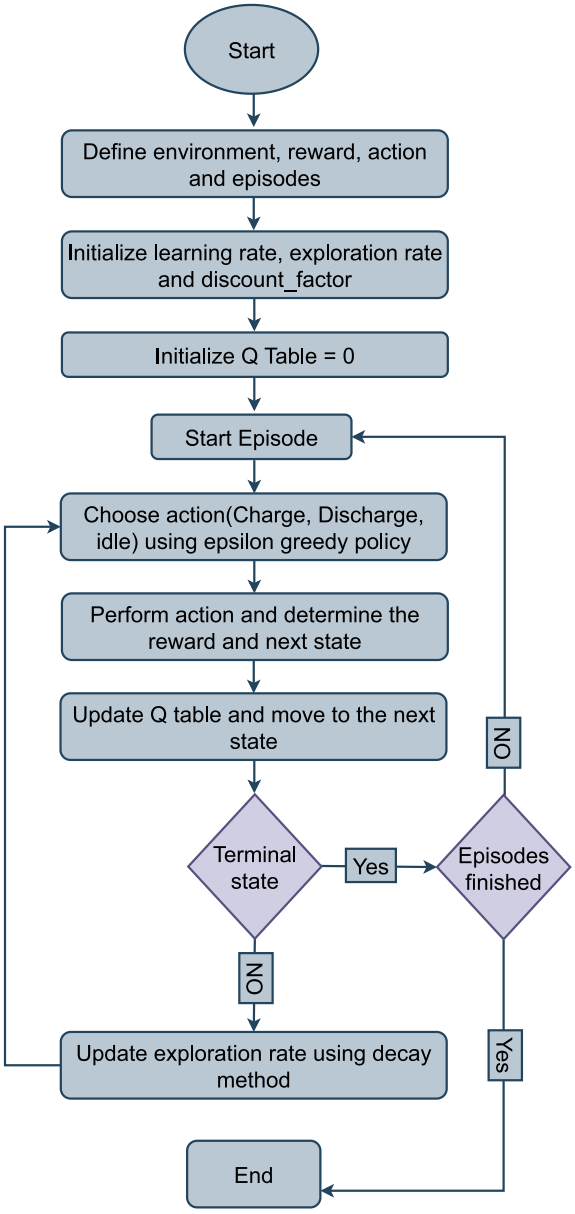


Fig. 4. Flow chart of the proposed algorithm.

3.5.1. State space (S)

This study incorporates two state variables, namely the time component denoted as *hour* and the battery charge component denoted as SOC. Eq. (5) illustrates the state space for the battery management environment.

$$S = \{hour, SOC\} \tag{5}$$

The temporal component *hour* represents the hour of day which allows the learning agent to learn about dairy farm load consumption and PV energy generation. SOC represents battery State of Charge (SOC) controllability. In this study, SOC was divided into ten bins, ranging from 0 to 9. Each bin corresponds to a 10% increment of the battery charge, effectively discretizing the state space of the battery management environment. This approach is taken to ensure that the distribution of SOC is evenly distributed and simplifies the complexity of the environment, making it more effective for analysis. SOC is represented as $SOC = SOC_c/SOC_{max}$. The SOC_c represents the battery charge at the current timestamp and SOC_{max} represents the battery maximum capacity.

3.5.2. Action space (A)

This study examines a set of three actions, namely charging, discharging, or remaining idle, represented as $A = \{charge, discharge, idle\}$, where an action $A = charge$, representing the charging of the battery using PV, and from the local utility grid. If $A = discharge$ discharge the battery when necessary to meet some or all of the energy requirements. In cases where the energy provided by the PV system and the battery is insufficient, it may be necessary to purchase additional power from the grid. If $A = idle$, the battery is in an idle state and the dairy farm is powered via solar PV and the grid. For selecting the action the Epsilon greedy policy is used.

In reinforcement learning, the epsilon-greedy policy is an approach that is often used with the Q-learning algorithm. It is the policy that helps the agent select an action in a specific state by using exploration and exploitation methods. In exploration, the agent chooses an action randomly without using previous knowledge of the environment, but in exploitation, the agent chooses the action using previous knowledge. The agent decides on the exploration based on the value of ϵ, which ranges from 0 to 1. If $\epsilon = 0.1$, then there is a 10% chance that the agent will explore the state and take random action on that state.

3.5.3. Reward (R)

The reward function, denoted as R, is computed as the cost of electricity imported from the grid and the electricity price at that hour. Eq. (6) represents the detailed mathematical formulation to calculate the reward for the battery management environment

$$R = \begin{cases} -((P_{dem} + (\beta - P_{pv})) \times P_e) - P_c \text{ if } A = charge \\ -(((P_{dem} - P_{pv}) - \gamma) \times P_e) - P_d \text{ if } A = discharge \\ -((P_{dem} - P_{pv}) \times P_e) - P_i \text{ if } A = idle \end{cases} \quad (6)$$

In Eq. (6) R represents the reward obtained at time t, P_{pv} denotes the total power generated by the solar panels at time t, P_{dem} represents the demand of the electricity by the dairy farm, β represent the charge rate at which battery is charged. γ represents the discharge rate at which the battery is discharged in kW, and A represents the action taken at time t, which can be either charge, discharge, or idle. P_e represents the price of electricity at the current time. The P_c represents the penalty amount by which the agent is penalized if it charges the battery under certain rules, P_d is the penalized amount when the agent selects an action to discharge the battery, and P_i is the amount of penalty when agent selects an action idle. The formulations of the penalized terms, which are applied based on the actions taken by the agent, are outlined in Eqs. (7), (8), and (9).

$$P_c = \begin{cases} -15 \text{ if } SOC_c \geq SOC_{\max} \text{ and hour == peak hours} \\ -10 \text{ if } SOC_c \geq SOC_{\max} \\ -10 \text{ if hour == peak hours} \\ +5 \text{ if hour == off-peak hours} \end{cases} \quad (7)$$

Eq. (7) explains how penalties are calculated when an agent chooses an action charge. This penalty depends on the battery's current state of charge (SOC_c) and the time of day. If the agent charges an already full battery (SOC_{max}), it gets penalized. A penalty of −15 is applied if the agent charges during peak electricity price hours and the battery is fully charged. The agent is penalized a penalty of −10 in two scenarios: first, if it charges the battery during off-peak hours when the battery is already fully charged, and second if it charges the battery during peak electricity hours. Contrarily, the agent gets a penalty of +5 for favorable actions like charging the battery at night when electricity prices are lower.

$$P_d = \begin{cases} -10 \text{ if } SOC_c \leq SOC_{\min} \\ -5 \text{ if hour == off-peak hours} \\ +5 \text{ if hour == peak hours and if } SOC_c > SOC_{\min} \end{cases} \quad (8)$$

Eq. (8) explains how penalties are calculated for discharging the battery. This penalty varies based on the battery's current state of charge (SOC_c) and the time of day. The agent faces a penalty of −10 if it discharges the battery below its minimum charge level (SOC_{min}). A penalty of −5 is applied if the battery is discharged during off-peak times. However, discharging during peak hours periods results in a reward of +5 if the battery charge is more than the battery's minimum level.

$$P_i = \begin{cases} -10 \text{ if } SOC_c \geq SOC_{\min} \text{ and hour == peak hours} \end{cases} \quad (9)$$

Eq. (9) outlines the penalty for the agent when it selects the "Idle" action. This penalty depends on the battery's current state of charge (SOC_c) and the time of day. To encourage more usage of battery power, a penalty of −10 is imposed during peak hours if the battery's charge level is above the minimum level.

The proposed Q-learning algorithm for battery management in dairy farming is presented in Algorithm 3.

Algorithm 3 Battery Management using Q-learning

Initialize $days$, $hours$, $maxSOC$, $learning\_rate$, $discount\_factor$, $epsilon$, $decay$, $steps\_per\_episode$, $total\_episodes$
Initialize $actions \leftarrow$ {'charge', 'discharge', 'idle'}
Initialize $Q\_table[hours][maxSOC + 1][\text{len}(actions)] \leftarrow 0$
for $episode = 1$ **to** $totalEpisodes$ **do**
    $hour \leftarrow 1$
    $SOC \leftarrow$ random between 1 and 10
    while $steps\_per\_episode$ **do**
        Choose $action$ from $actions$ using ϵ-greedy policy
        Take $action$, observe $reward$, new_hour, new_SOC
        $Q\_value \leftarrow Q\_table[hour][SOC][\text{index of } action]$
        $next\_Q\_value \leftarrow \max(Q\_table[new\_hour][new\_SOC])$
        $Q\_table[hour][SOC][\text{index of } action] \leftarrow Q\_value + learning\_rate \times (reward + discount\_factor \times (next\_Q\_value - Q\_value))$
        $hour, SOC \leftarrow new\_hour, new\_SOC$
    end while
    $learning\_rate \leftarrow \max(learning\_rate - decay, 0.1)$
    $epsilon \leftarrow \max(epsilon - decay, 0.1)$
end for

Algorithm 3 describes a Q-learning method specifically designed for battery management in dairy farming. It initializes Q-values for each state–action pair and then iterates through one million episodes. Within each episode, the algorithm selects an action based on a policy derived from the Q-values, such as the ϵ-greedy strategy. After choosing an action, it observes the reward and the next state that results from that action. The algorithm then updates the Q-value for the current state–action pair. Then algorithm uses the weight decay method to decrease the exploration and learning rate with respect to the number of episodes and set it to a minimum of 0.1. Finally, it repeats the process for all episodes.

3.6. Experimental setup

This research evaluates the proposed Q-learning algorithm for battery management through a series of experiments.

1. Experiment 1 involves testing and training the Q-learning algorithm on the Finland dairy farm electricity data.
2. Experiment 2 incorporates Finland wind data for a more detailed evaluation of the algorithm.
3. Experiment 3 tests the performance of the algorithm by exploring the state space.
4. Experiment 4 applies the algorithm to the Irish dairy farm data.

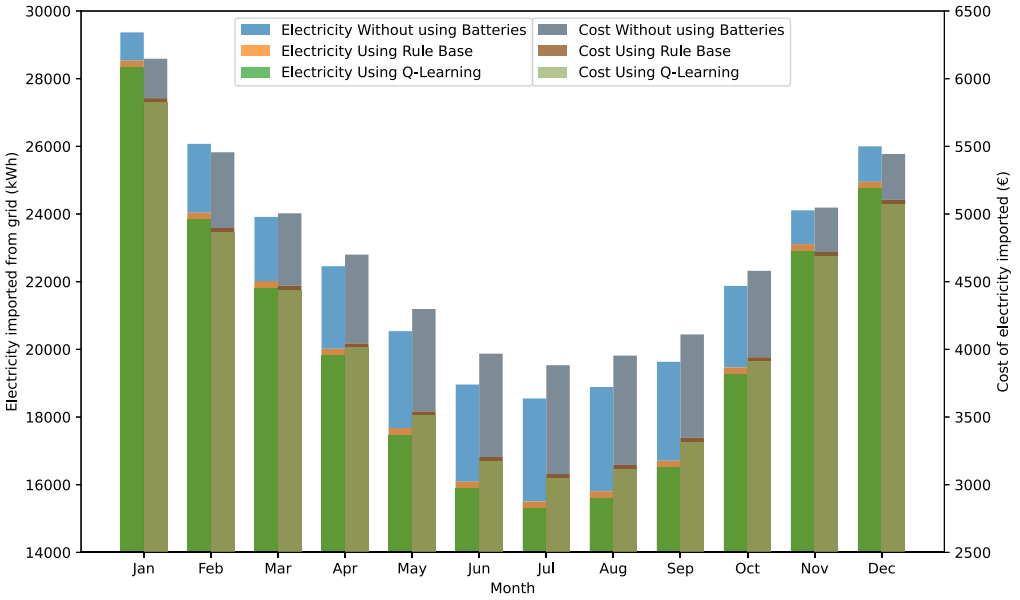


Fig. 5. Comparison of the electricity load and cost imported from the power grid by using rule-base and Q-learning on Finland dataset.

The goal of these experiments is to assess the algorithm's effectiveness in different scenarios, involving parameter adjustments, data analysis, and comparative studies. These experiments aim to demonstrate the algorithm's robustness and potential for optimizing dairy farm energy use.

4. Results and discussion

4.1. Q-learning for battery management

In this scenario, the Q-learning algorithm was trained to enhance the efficiency of battery management in dairy farming. Its primary goal is to increase the use of PV energy while reducing dependence on the external power grid and to lower energy cost in the dairy farm. This algorithm was trained on one year's data from Finland [43]. The trained algorithm learned the optimal policy for charging the battery, discharging, and remaining idle, considering state information on battery charge level, time, and energy prices. After training, the algorithm's performance was tested on the same dataset for one year. The findings indicate that the implementation of the Q-learning algorithm decreases the import of electricity by 10.64%. In comparison, the baseline strategy resulted in a decrease in electricity imports only by 9.72%. This improvement in the reduction of electricity imports from the grid is presented in Fig. 5, demonstrates the algorithm's effectiveness.

Fig. 5 shows a comparison of total electricity imported from the grid and the associated cost of the electricity in each month of the year. The x-axis shows the time in months, while the y-axis on the left side indicates the total electricity imported from the grid while the y-axis on the right side demonstrates the cost of the electricity imported. The graph shows two distinct bars representing: electricity imported from the grid using three methods each marked with a different color; and the cost of the imported electricity by comparing it with three methods, each depicted in a different color. This illustration offers a clear insight into how different energy management strategies affect the overall consumption of electricity and reliance on the grid. The Q-learning effectively reduced electricity imports by 10.64% and cost by 13.41% as compared to the baseline algorithm which reduces electricity import by 9.72% and cost by 12.73%. These results highlight the effectiveness of Q-learning in optimizing energy usage compared to rule-based battery management and without battery management in reducing grid dependency.

Fig. 6 illustrates a comparison of battery charging behaviors throughout a day using two methodologies: baseline and Q-learning. Additionally, it displays electricity price, consumption, PV, and wind generation data for the first day of the year. The x-axis represents the hours of the day, while the y-axis indicates the battery and electricity profiles on the farm. This comparison highlights differences in battery charging and discharging behaviors between the two methodologies. The Q-learning method demonstrates enhanced battery management, with results indicating an optimal policy for charging and discharging. Specifically, when it charges the battery during periods of low electricity prices and available PV and wind generation, maximizing the utilization of renewable energy sources. Conversely, during peak hours when electricity prices are high, the Q-learning algorithm discharges the battery. In contrast, the rule-based method follows a more static approach, based on predetermined rules while Q-learning is adaptive to the current environment. This adaptability allows for more effective optimization of the battery charging and discharging, aligning with fluctuating energy demands and variable PV and wind generation, leading to enhanced efficiency and cost savings for the dairy farm.

The peak demand metric is calculated to determine the benefit of Q-learning in terms of its impact on the grid. The algorithm achieved a 2% reduction in peak demand when using battery management, which is crucial for reducing load from the power grid and reducing the electricity cost in the dairy farm. This reduction, illustrated in Fig. 7, compares the peak demand load imported from the grid using Q-learning and without battery management in the month for 1 year, emphasizing the algorithm's effectiveness during periods of peak demand. This significant reduction is particularly important for practical energy management to reduce electricity demand during periods of peak demand.

4.2. Battery management with wind generation

This study investigates the impact of wind energy on the efficacy of the Q-learning algorithm, utilizing the Finland dataset which captures wind generation metrics. The Q-learning algorithm was trained for a total of one million episodes utilizing wind data, in addition to solar data and a load demand from a farm over one year. After training the algorithm performance is evaluated on the data. The objective of this experiment is to assess the efficacy of Q-learning in energy management by incorporating both wind and solar sources.

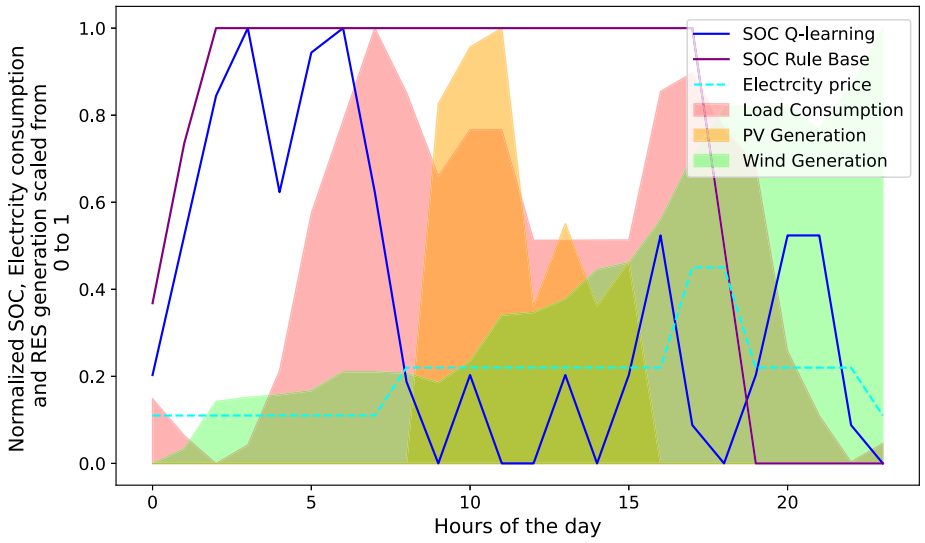


Fig. 6. Comparison of the battery charging and discharging by using TOU and Q-learning.

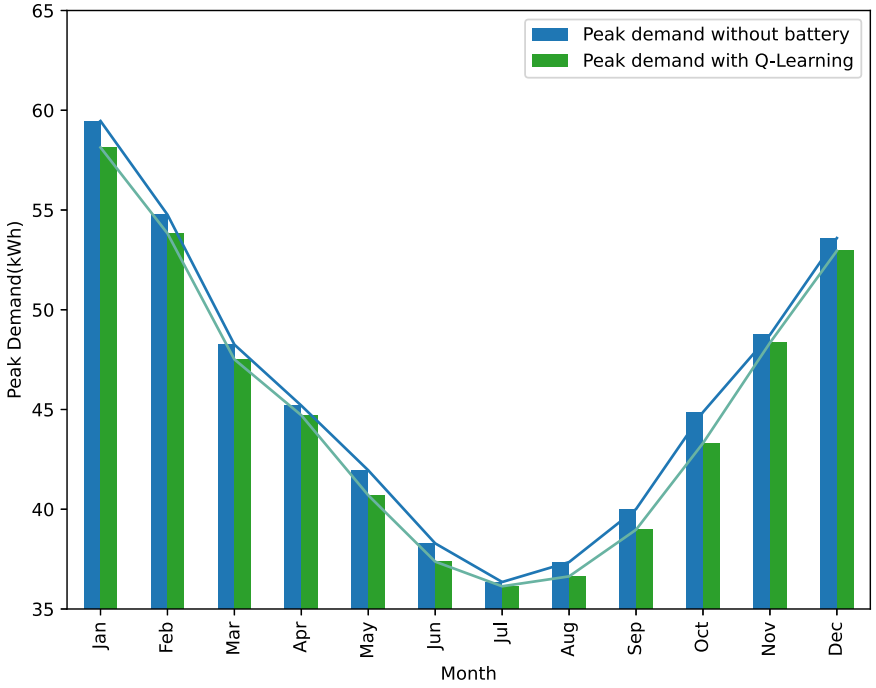


Fig. 7. Comparison of the reduction of the peak demand on the grid.

Fig. 8, shows a comparison between the electricity imported from the grid by utilizing wind energy and without wind energy. The x-axis of the figure represents the months of the year, while the vertical axis represents the electricity imported from the grid. The figure shows that the utilization of wind energy resulted in a decrease of 22.14% in the import of grid electricity, in comparison to 10.64% generated without wind energy.

The findings of the experiment demonstrate that incorporating wind energy through the utilization of the Q-learning algorithm leads to significant reductions in the cost of imported electricity. The reductions using wind energy reduce electricity cost by 24.49% compared to 13.41% reduced without wind energy. By integrating wind energy, the algorithm comprehensively reduces electricity import from the grid during the winter period because the wind generation is high in comparison to the PV generation due to wind storms. In summer periods the wind is not too high which affects the performance of the algorithm. The above-mentioned results show the efficiency of the Q-learning algorithm for battery management and the reduction of imported electricity from the grid when wind energy is incorporated.

4.3. Investigating state space

In this experiment, we explore the impact of expanding the state space on the performance of the Q-learning algorithm, initially developed in Experiment 4.1. The state space of the first experiment is

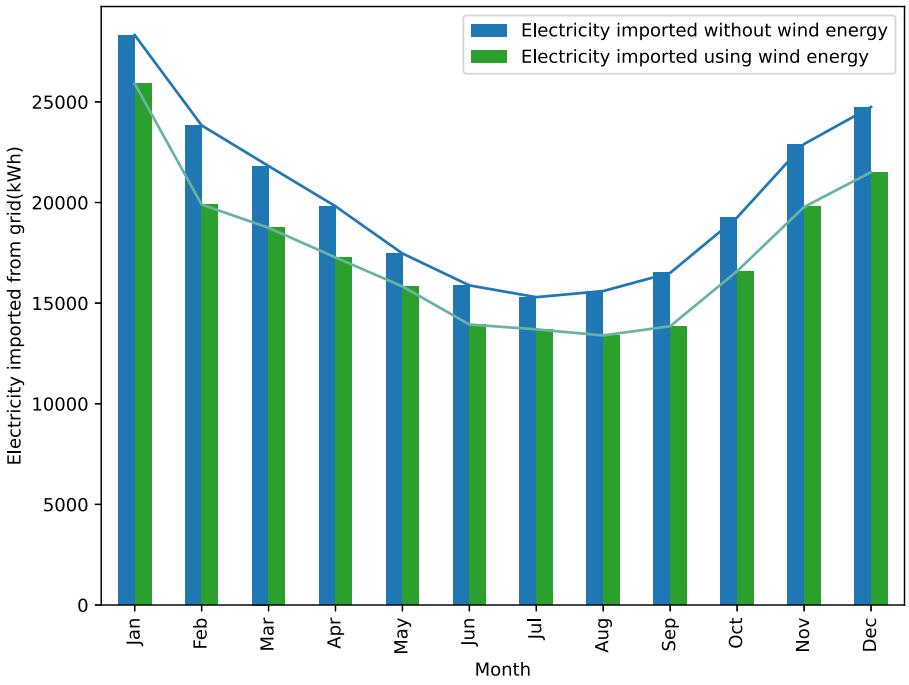
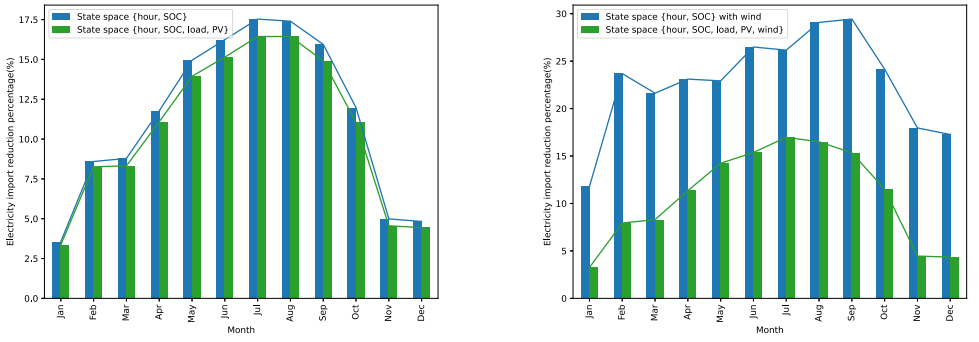


Fig. 8. Comparison of the electricity load imported from the power grid by using Q-learning without wind and with wind energy.



(a) Electricity import reduction without wind (b) Electricity import reduction with wind

Fig. 9. Comparison of electricity import reduction percentage with different state space.

depicted in Eq. (5). The load demand and PV generation were used in the reward function to calculate the reward.

In this investigation, the load and PV generation are incorporated into the state space of the Q-learning algorithm. The purpose of this is to observe how the algorithm's performance is affected when these variables are part of the state space, instead of using them to calculate reward. The formulation of this modification is depicted in Eq. (10).

$$S = \{hour, SOC, load, PV\} \tag{10}$$

To further explore the algorithm's adaptability and efficiency, we expanded the state space to include wind data information. Exploration aims to see how dynamic state space affects the adaptability and efficiency of the algorithm. Also, to see how dynamic state space affects the algorithm learning and decision-making capabilities when wind generation data is added to the state space for the battery management system. The state space for this extended approach, incorporating wind data, is presented in Eq. (11).

$$S = \{hour, SOC, load, PV, wind\} \tag{11}$$

The Q-learning algorithm is trained and tested with different state spaces including scenarios with and without wind generation data. Fig. 9 compares the algorithm's performance across these different state spaces. Fig. 9(a) shows how the inclusion of load and PV generation in the state space affects electricity import reduction, compared to the state space from experiment 4.1. We found that the state space from experiment 4.1 is more effective, reducing electricity imports by 10.64%, compared to the modified state space (with load, and PV) which only achieved a 9.97% reduction. Fig. 9(b) illustrates the impact of incorporating wind generation data into the state space. When the state space from Experiment 4.1 is combined with load demand, PV generation and wind data, then there is a smaller reduction in load import, achieving a 10.07% decrease. In contrast, state space from Experiment 4.1 with wind generation data results in a significant reduction of 22.14%. This shows that expanding the state space adds to the complexity of the environment, which makes it difficult for the Q-learning agent to make optimal decisions. Another reason for this incapability could be due to the curse of dimensionality in Q-learning, where increasing dimensionality leads to sparser data and challenges in achieving expected results [50]. Additionally, discretizing the state space to manage its dimensionality might have affected performance due to variations in the data.

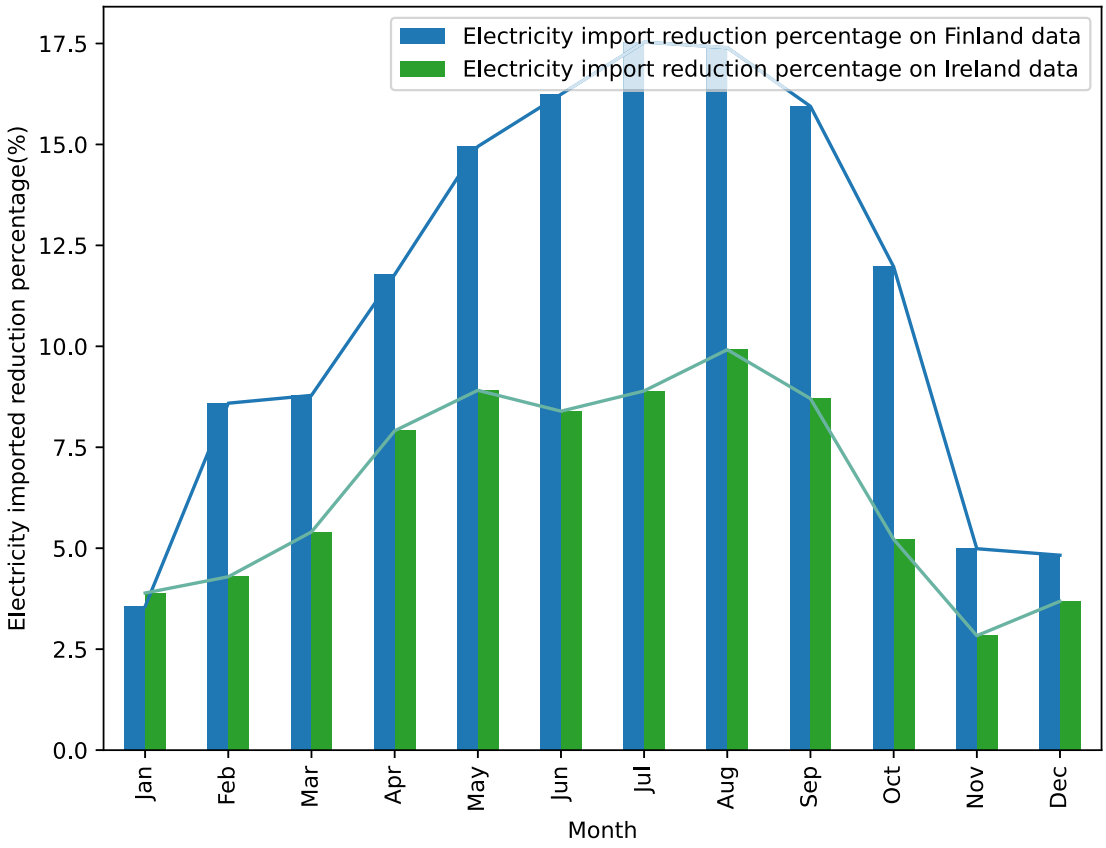


Fig. 10. Comparison of the electricity import reduction percentage on Finland and Ireland data using Q-learning.

4.4. Irish dairy farm case study

In this study, we applied the Q-learning algorithm, originally developed in Experiment 4.1, to the context of Irish dairy farms. The primary objective was to test the algorithm's adaptability using a dataset collected specifically for Ireland. We focused on analyzing electricity consumption and PV energy generation patterns. The main goal of this experiment was to evaluate the efficacy of the Q-learning algorithm in adapting to new data patterns, aiming to optimize battery scheduling and decrease reliance on the electricity grid.

The comparison of the percentage of electrical load imported from the grid using Q-learning, based on datasets from Finland and Ireland, is illustrated in Fig. 10. The figure illustrates that the algorithm shows better results in reducing electricity import percentages when applied to the Finland data as compared to the Ireland data. This difference is because the algorithm was trained on the Finland dataset, allowing it to learn and adapt to its specific patterns of electricity consumption and PV generation. In contrast, the Ireland dataset represents a new environment with variations in consumption and generation patterns, which is a new environment for the algorithm in exploring states and deciding on charging and discharging actions. To provide a comprehensive overview of the results of this experiment, we have detailed the results for both Ireland and Finland data in Table 1, comparing the proposed algorithm with a baseline algorithm.

Table 1 shows a comparison of the performance between the baseline method and the Q-learning approach. It highlights the Q-learning algorithm's capability in effectively lowering imported grid load and related cost. Specifically, the Q-learning algorithm reduces electricity import on Ireland data by 6.7%, an improvement over the baseline's algorithm which reduces by 5.54%. Additionally, the cost associated with the load was reduced by up to 9.37% in comparison with the baseline which was reduced by 8.50%. This comparison showed the adaptability of the Q-learning algorithm in optimizing electricity load and the cost associated with it.

Table 1
Comparison of load and cost reductions for Q-learning and Rule-Base algorithms on Finland and Ireland datasets.

Country	Rule-based		Q-learning	
	Load(%)	Cost(%)	Load(%)	Cost(%)
Finland	9.72	12.73	10.64	13.41
Ireland	5.54	8.50	6.70	9.37

5. Conclusion

In this research, Q-learning is applied to battery management in a dairy farm, using electricity data from Finland. This study involved various experiments to assess the effectiveness of the Q-learning algorithm. This research explored the effect of integrating wind and solar data on battery management and examined how changing the state space of the algorithm impacts its performance. Additional experiments were conducted using data from Ireland to validate the effectiveness of the algorithm. As explained in Section 4, the findings show that the Q-learning algorithm successfully reduced the reliance of the dairy farm on the external grid.

Below are the main findings of this research:

1. This research utilized Q-learning to manage battery energy in dairy farms, resulting in efficient scheduling of battery loads. The implementation of this strategy resulted in a significant decrease of 13.41% in the cost of electricity imported from the grid and a reduction in peak demand of 2%. This shows the proposed strategy's potential to address energy management within the context of dairy farming effectively.

2. The Q-learning algorithm, when applied to wind data integrated with solar data, demonstrated impressive results, achieving a substantial reduction in imported electricity cost by 24.49%. This emphasizes the algorithm's effectiveness in managing batteries efficiently when wind-generated energy is incorporated with solar energy.
3. Exploring different state spaces in the Q-learning algorithm led to a reduction in electricity import cost. Different experiments were conducted by expanding state space to see the expandability and adaptability of the algorithm. This improvement highlights the impact of modifying state spaces on battery management in dairy farming when using a Q-learning algorithm.
4. Testing the Q-learning algorithm on the Ireland dataset significantly decreased electricity imports from the grid, with a notable reduction of 6.7% compared to the 5.54% achieved with the baseline approach. The outcome shows the Q-learning algorithm's adaptability and effectiveness when applied to data from various regions.

In the future, we intend to employ DRL algorithms to address the challenge of state space expansion. Deep Learning techniques are well-suited for handling complex problems, and by integrating them, we aim to enhance the model's ability to handle complex state space. This strategy will enhance performance by decreasing dependence on the external grid.

CRediT authorship contribution statement

Nawazish Ali: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Abdul Wahid:** Writing – review & editing, Supervision, Project administration, Investigation, Formal analysis. **Rachael Shaw:** Writing – review & editing, Validation, Supervision, Investigation, Formal analysis. **Karl Mason:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Karl Mason reports financial support was provided by Science Foundation Ireland. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number [21/FFP-A/9040].

References

[1] U. Nations, Food and agriculture organization of the united nations, 2022, URL https://www.fao.org/home/en/. (Online; Accessed 20 April 2023).
[2] OECD, Dairy and dairy products, 2020, https://www.oecd-ilibrary.org/sites/aa3fa6a0-en/index.html?itemId=/content/component/aa3fa6a0-en. (Accessed 27 November 2022).
[3] J. Upton, M. Murphy, P. French, P. Dillon, Dairy farm energy consumption, in: Dairying: Entering a Decade of Opportunity. Teagasc National Dairy Conference 2010, 2010, pp. 87–97.
[4] Renewable energy opportunities for dairy farmers, 2021, URL https://ahdb.org.uk/knowledge-library/renewable-energy-opportunities-for-dairy-farmers. (Online; Accessed 20 April 2023).
[5] Energy, 2020, https://www.gov.ie/en/policy/9cd812-energy/. (Accessed 26 June 2023).
[6] U.E.I. Administration, Electricity in the U.S., 2022, https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php. (Accessed 26 June 2023).
[7] M. Hannan, S. Wali, P. Ker, M. Abd Rahman, M. Mansor, V. Ramachandaramurthy, K. Muttaqi, T. Mahlia, Z. Dong, Battery energy-storage system: A review of technologies, optimization objectives, constraints, approaches, and outstanding issues, J. Energy Storage 42 (2021) 103023.
[8] B. Zou, J. Peng, S. Li, Y. Li, J. Yan, H. Yang, Comparative study of the dynamic programming-based and rule-based operation strategies for grid-connected PV-battery systems of office buildings, Appl. Energy 305 (2022) 117875.
[9] B.M. Kumar, V. Talukdar, H. Khan, S.B. Talukdar, A. Koujalagi, R.G. Kumar, A. Gupta, 6 application of al-based, in: Robotics and Automation in Industry 4.0: Smart Industries and Intelligent Technologies, CRC Press, 2024, p. 110.
[10] V. Mnih, A.P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, K. Kavukcuoglu, Asynchronous methods for deep reinforcement learning, in: International Conference on Machine Learning, PMLR, 2016, pp. 1928–1937.
[11] C.J. Watkins, P. Dayan, Q-learning, Mach. Learn. 8 (1992) 279–292.
[12] D. Azuatalam, K. Paridari, Y. Ma, M. Förstl, A.C. Chapman, G. Verbič, Energy management of small-scale PV-battery systems: A systematic review considering practical implementation, computational requirements, quality of input data and battery degradation, Renew. Sustain. Energy Rev. 112 (2019) 555–570.
[13] Y. Zhang, T. Ma, P.E. Campana, Y. Yamaguchi, Y. Dai, A techno-economic sizing method for grid-connected household photovoltaic battery systems, Appl. Energy 269 (2020) 115106.
[14] M. Braun, K. Büdenbender, D. Magnor, A. Jossen, Photovoltaic self-consumption in Germany: using lithium-ion storage to increase self-consumed photovoltaic energy, in: 24th European Photovoltaic Solar Energy Conference (PVSEC), Hamburg, Germany, 2009.
[15] D. Talavera, F. Muñoz-Rodriguez, G. Jimenez-Castillo, C. Rus-Casas, A new approach to sizing the photovoltaic generator in self-consumption systems based on cost–competitiveness, maximizing direct self-consumption, Renew. Energy 130 (2019) 1021–1035.
[16] N.J. Vickers, Animal communication: when i'm calling you, will you answer too? Curr. Biol. 27 (14) (2017) R713–R715.
[17] R. Luthander, J. Widén, D. Nilsson, J. Palm, Photovoltaic self-consumption in buildings: A review, Appl. Energy 142 (2015) 80–94.
[18] V. Sharma, M.H. Haque, S.M. Aziz, Energy cost minimization for net zero energy homes through optimal sizing of battery storage system, Renew. Energy 141 (2019) 278–286.
[19] E. Nyholm, J. Goop, M. Odenberger, F. Johnsson, Solar photovoltaic-battery systems in Swedish households–self-consumption and self-sufficiency, Appl. Energy 183 (2016) 148–159.
[20] L. Dusonchet, E. Telaretti, Comparative economic analysis of support policies for solar PV in the most representative EU countries, Renew. Sustain. Energy Rev. 42 (2015) 986–998.
[21] C.M. Flath, An optimization approach for the design of time-of-use rates, in: IECON 2013-39th Annual Conference of the IEEE Industrial Electronics Society, IEEE, 2013, pp. 4727–4732.
[22] R. Li, Z. Wang, C. Gu, F. Li, H. Wu, A novel time-of-use tariff design based on Gaussian mixture model, Appl. Energy 162 (2016) 1530–1536.
[23] N.R. Darghouth, R.H. Wiser, G. Barbose, Customer economics of residential photovoltaic systems: Sensitivities to changes in wholesale market design and rate structures, Renew. Sustain. Energy Rev. 54 (2016) 1459–1469.
[24] M. Gitizadeh, H. Fakharzadegan, Battery capacity determination with respect to optimized energy dispatch schedule in grid-connected photovoltaic (PV) systems, Energy 65 (2014) 665–674.
[25] A.S. Hassan, L. Cipcigan, N. Jenkins, Optimal battery storage operation for PV systems with tariff incentives, Appl. Energy 203 (2017) 422–441.
[26] E.L. Ratnam, S.R. Weller, C.M. Kellett, An optimization-based approach to scheduling residential battery storage with solar PV: Assessing customer benefit, Renew. Energy 75 (2015) 123–134.
[27] Q. Wei, D. Liu, G. Shi, A novel dual iterative Q-learning method for optimal battery management in smart residential environments, IEEE Trans. Ind. Electron. 62 (4) (2014) 2509–2518.
[28] S. Kim, H. Lim, Reinforcement learning based energy management algorithm for smart energy buildings, Energies 11 (8) (2018) 2010.
[29] F. Ruelens, B.J. Claessens, S. Quaiyum, B. De Schutter, R. Babuška, R. Belmans, Reinforcement learning applied to an electric water heater: From theory to practice, IEEE Trans. Smart Grid 9 (4) (2016) 3792–3800.
[30] B. Li, L. Xia, A multi-grid reinforcement learning method for energy conservation and comfort of HVAC in buildings, in: 2015 IEEE International Conference on Automation Science and Engineering, CASE, IEEE, 2015, pp. 444–449.
[31] E. Foruzan, L.-K. Soh, S. Asgarpoor, Reinforcement learning approach for optimal distributed energy management in a microgrid, IEEE Trans. Power Syst. 33 (5) (2018) 5749–5758.
[32] C. Guan, Y. Wang, X. Lin, S. Nazarian, M. Pedram, Reinforcement learning-based control of residential energy storage systems for electric bill minimization, in: 2015 12th Annual IEEE Consumer Communications and Networking Conference, CCNC, IEEE, 2015, pp. 637–642.

[33] Y. Liu, D. Zhang, H.B. Gooi, Optimization strategy based on deep reinforcement learning for home energy management, CSEE J. Power Energy Syst. 6 (3) (2020) 572–582.
[34] J. Cao, D. Harrold, Z. Fan, T. Morstyn, D. Healey, K. Li, Deep reinforcement learning-based energy storage arbitrage with accurate lithium-ion battery degradation model, IEEE Trans. Smart Grid 11 (5) (2020) 4513–4521.
[35] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, T. Jiang, Deep reinforcement learning for smart home energy management, IEEE Internet Things J. 7 (4) (2019) 2751–2762.
[36] S. Abedi, S.W. Yoon, S. Kwon, Battery energy storage control using a reinforcement learning approach with cyclic time-dependent markov process, Int. J. Electr. Power Energy Syst. 134 (2022) 107368.
[37] Z. Wei, Z. Quan, J. Wu, Y. Li, J. Pou, H. Zhong, Deep deterministic policy gradient-DRL enabled multiphysics-constrained fast charging of lithium-ion battery, IEEE Trans. Ind. Electron. 69 (3) (2021) 2588–2598.
[38] B. Huang, J. Wang, Deep-reinforcement-learning-based capacity scheduling for PV-battery storage system, IEEE Trans. Smart Grid 12 (3) (2020) 2272–2283.
[39] G. Cheng, L. Dong, X. Yuan, C. Sun, Reinforcement learning-based scheduling of multi-battery energy storage system, J. Syst. Eng. Electron. 34 (1) (2023) 117–128.
[40] D. Paudel, T.K. Das, A deep reinforcement learning approach for power management of battery-assisted fast-charging EV hubs participating in day-ahead and real-time electricity markets, Energy 283 (2023) 129097.
[41] Tesla.com, How powerwall works, 2023, https://www.tesla.com/support/energy/powerwall/learn/how-powerwall-works. (Online; Accessed 27 March 2023).
[42] M. Hannan, S. Wali, P. Ker, M. Abd Rahman, M. Mansor, V. Ramachandaramurthy, K. Muttaqi, T. Mahlia, Z. Dong, Battery energy-storage system: A review of technologies, optimization objectives, constraints, approaches, and outstanding issues, J. Energy Storage 42 (2021) 103023.
[43] S. Uski, E. Rinne, Data for a dairy farm microgrid solution, 2018, URL https://zenodo.org/record/1294967#.ZFD0Fc7MIQ8.
[44] N.R.E.L. (NREL), System advisor model (SAM), 2017, https://sam.nrel.gov. (Online; Accessed 1 November 2022).
[45] Electricity products and prices | Helen, 2023, URL https://www.helen.fi/en/electricity/electricity-products-and-prices. (Accessed 15 November 2023).
[46] Electric Ireland, Time-of-use tariffs for residential customers, 2022, URL https://www.electricireland.ie/residential/help/smart-electricity-meters/time-of-use-tariffs-for-residential-customers. (Online; Accessed 15 November 2022).
[47] H. Khaleghy, A. Wahid, E. Clifford, K. Mason, Modelling electricity consumption in irish dairy farms using agent-based modelling, in: Proceedings of the Artificial Intelligence for Sustainability Workshop (AI4S) at ECAI, 2023.
[48] Electric Ireland, New customer price plans, 2022, URL https://www.electricireland.ie/switch/new-customer/price-plans?priceType=P. (Online; Accessed 15 November 2022).
[49] R. Bellman, Dynamic programming, Science 153 (3731) (1966) 34–37.
[50] R. Sutton, Barto: "reinforcement learning: An introduction", IEEE Trans. Neural Netw. 9 (1998) 1054.