Control Strategy of a Hybrid Renewable Energy System Based on Reinforcement Learning Approach for an Isolated Microgrid

Bao Chau Phan and Ying-Chih Lai *

Department of Aeronautics and Aeronautics, National Cheng Kung University, Tainan 701, Taiwan; pbchau.hk09@gmail.com
* Correspondence: yingclai@mail.ncku.edu.tw; Tel.: +886-6-275-7575 (ext. 63648)

check for updates

Featured Application: This study has demonstrated an efficient maximum power point tracking method based on reinforcement learning to improve the renewable energy conversion. The theory can also be applied for the problems of the optimal sizing and energy management systems to develop the cost-efficient and environmentally friendly microgrids, especially for rural and islanding electrification.

Abstract: Due to the rising cost of fossil fuels and environmental pollution, renewable energy (RE) resources are currently being used as alternatives. To reduce the high dependence of RE resources on the change of weather conditions, a hybrid renewable energy system (HRES) is introduced in this research, especially for an isolated microgrid. In HRES, solar and wind energies are the primary energy resources while the battery and fuel cells (FCs) are considered as the storage systems that supply energy in case of insufficiency. Moreover, a diesel generator is adopted as a back-up system to fulfill the load demand in the event of a power shortage. This study focuses on the development of HRES with the combination of battery and hydrogen FCs. Three major parts were considered including optimal sizing, maximum power point tracking (MPPT) control, and the energy management system (EMS). Recent developments and achievements in the fields of machine learning (ML) and reinforcement learning (RL) have led to new challenges and opportunities for HRES development. Firstly, the optimal sizing of the hybrid renewable hydrogen energy system was defined based on the Hybrid Optimization Model for Multiple Energy Resources (HOMER) software for the case study in an island in the Philippines. According to the assessment of EMS and MPPT control of HRES, it can be concluded that RL is one of the most emerging optimal control solutions. Finally, a hybrid perturbation and observation (P&O) and Q-learning (h-POQL) MPPT was proposed for a photovoltaic (PV) system. It was conducted and validated through the simulation in MATLAB/Simulink. The results show that it showed better performance in comparison to the P&O method.

Keywords: HRES; optimal sizing; MPPT control; EMS and reinforcement learning

1. Introduction

Energy plays an important role in modern human life and the economic development of a country. Currently, fossil fuels are the main and most reliable form of energy resources for power generation to cater for the huge increase in energy demand around the world. Due to the rising cost of fossil fuels and environmental pollution, renewable energy resources such as solar, wind, biomass, geothermal, etc., have been recently considered as alternative resources for sustainable development. Most countries

in the Association of Southeast Asian Nations (ASEAN), especially Vietnam, Thailand, Indonesia, Malaysia, and the Philippines have recently begun paying attention to green energy and have become the most successful countries for renewable energy deployment [1].

With the cost reduction and technological improvement, renewable energy (RE) resources are being combined with conventional generator and storage systems to supply the load demand with low power generation cost, high efficiency and reliability, and low fuel consumption that can reduce the environmental pollution problem. Moreover, the standalone hybrid renewable energy system (HRES) for rural and islanding electrification could be more cost-effective than grid extension, which is estimated to cost US$10,000 to US$50,000 per kilometer [2]. The developed system for this project adopted the RE resources (solar and wind energies) as primary energy resources. In addition, an electrolyzer was applied to produce hydrogen that was contained in the hydrogen tank for the operation of fuel cells (FCs). The battery and FCs were used as the storage systems that supply energy in case of insufficiency, and a diesel generator functioned as a back-up system to fulfill the load in the event of bad weather conditions [3]. In HRES, FC can be used as an option for long term energy storage [4]. However, the slow dynamics of FC and its degradation for the frequent start-up and shut down cycles are major disadvantages. Therefore, the battery is introduced to such hybrid systems for taking care of power deficits and acting as a short term energy storage medium [5]. The combination of FC and battery along with photovoltaic (PV) and wind turbines (WTs) ensures an uninterrupted power supply to the load.

Based on the requirement of energy demand and various technologies of RE resources, it is important to figure out the optimal configuration or suitable sizing of hybrid energy system components which can decrease the system cost and retain high reliability. There are many methodologies applied to proper unit sizing of HRES components such as artificial intelligence [6,7], multi-objective design [8], an iterative technique [9], probabilistic approach [6], etc. It was summarized that the sizing methods based on AI and multi-objective design are one of the most potent and powerful tools [3]. In this project, the Hybrid Optimization Model for Multiple Energy Resources (HOMER) was used for designing and determining the optimal sizing of the HRES for the case study due to it being user-friendly and easily implemented.

The control strategy of HRES is necessary for improving the productivity and reliable operation of a power system working under the uncertainties of the RE resources and the dynamic loads. Under this system, maximum power point tracking (MPPT) control [10] is used to improve the conversion efficiency of solar and wind energy systems while an energy management system (EMS) [11] is developed for controlling the power flows among system components and handling reliable operation. The recent development and achievement in the fields of machine learning (ML) and reinforcement learning (RL) leads to new challenges and opportunities for energy management and MPPT control. RL-based control systems can learn and act just following experiences when interacting with the environment [12,13]. In contrast, traditional methods need particular mathematical models of the system and environment, which requires highly control knowledge, data, and domain expertise. ML can be classified into three categories: supervised learning (task-driven, estimate next value), unsupervised (data-driven, determine clusters), and reinforcement learning (learning from trial and error) [13]. The RL controller can be considered as an agent, and the agent can learn how to act based on the reward and current states received from the environment [12]. Due to the potential development of ML in various areas, several researchers have switched their interest in the application of ML towards the control strategies of HRES, which will be discussed in the following sections.

This study aims to generally present the overall process of energy system development for an isolated microgrid, especially for a hybrid renewable hydrogen energy system involving optimal sizing, EMS, and an MPPT controller. Firstly, HOMER software was used for the optimal sizing of HRES based on the actual load demand and weather data in Basco Island, Bantanes, Philippines as the case study. Next, a brief review of EMS and MPPT control based on the ML and RL techniques was conducted. Finally, a new hybrid method for MPPT control, which integrates the Q-learning

and perturbation and observation (P&O) methods, was proposed to improve system performance. P&O is the mostly preferred algorithm for MPPT control [14,15]. The major advantages of this method are simple structure and ease of implementation. But P&O turned out to be ineffective under the fast change of the temperature and irradiation, as well as the partial shading conditions. A large step size of the P&O duty cycle (D) provides fast convergence with poor tracking while the low step size duty cycle provides low convergence with the ability to reduce the oscillation at the maximum power point [15]. The reinforcement learning approach to solve the MPPT problem aims to learn the system behavior based on the PV source response. The RL-based MPPT controller monitors the environmental state of the PV source and uses different step sizes of the duty cycle to adjust the perturbation to the operating voltage in achieving the maximum power. In references [16,17], the authors present the good simulation results of reinforcement learning. In addition, reference [18] has presented considerable results towards a universal MPPT control method. It has also mentioned the potential future research on this topic including state-space reduction, RL algorithm optimization, a comparison between different RL algorithms, a more efficient optimal procedure, and the practical experiments. As discussed, we aim to combine the Q-learning with the P&O method to reduce the state space for the learning process and to enhance the good characteristics of the P&O controller in this study.

The major contributions of this study are as follows:

- The optimal sizing of hybrid renewable hydrogen energy system by HOMER was presented for the case study based in Basco island, The Philippines.
- A proposed robust MPPT control based on the Q-learning and P&O methods, named as h-POQL, was simulated and validated in MATLAB/Simulink.
- The simulation of the proposed h-POQL shows that the P&O controller can tune the reference input values of the duty cycle and track the maximum power point with faster speed and high accuracy based on the optimal results learned by the Q-learning algorithm.
- A comparison between the h-POQL and the P&O method was carried out.

This paper is organized as follows. Section 2 presents the review of the energy management systems of HRES based on RL. Section 3 shows the optimal sizing of HRES based on HOMER software. A quick review of MPPT control methods and the proposed h-POQL controller was conducted in Section 4. Finally, the discussions were presented in Section 5, while Section 6 provide the conclusions and future work areas.

2. The Assessment of the Energy Management System for HRES

The literature survey on EMS shows that related studies are quite extensive and consists of various hybrid system configurations [4,19]. The energy management strategies are usually dependent on the type of energy system, including standalone, grid-connected, and smart grid as mentioned in reference [11]. Besides, the EMS architectures can be classified into three groups: centralized, distributed, and hybrid centralized and distributed controllers [20]. The advantage of centralized control is that it can handle the problems of multi-objective energy management and obtain the global optimal solution, while the distributed controller can reduce the computational time and detect the single-point failure.

In general, the control strategies can be divided into two categories: classical and intelligent control. Some EMS studies are based on classical techniques, such as linear and nonlinear programming, dynamic programming, ruled-based and flowchart methods [11]. In addition, the proportional integral (PI) controllers and some nonlinear controllers such as sliding mode controller and H-infinity controller are presented in reference [21]. The advantage of these controllers is that they require a low computational burden. However, the implementation and tuning would be more complicated due to the increase in the number of variables. It is not easy to obtain the mathematical model of HRES based on these techniques, and they are also heavily dependent on complex mathematical analysis.

Due to the drawbacks of the conventional-based EMS methods, intelligent control strategies, which are more robust and efficient, have been developed, such as fuzzy logic control (FLC) [22], an artificial neural network (ANN), an adaptive neuro-fuzzy inference system (ANFIS) [23], a model predictive controller (MPC), etc. [20]. Moreover, evolutionary algorithms, such as Particle Swarm Optimization (PSO) and the Genetic Algorithm [20], have been studied to optimize the controllers used for solving the multi-objective optimization problem. In addition, research on the prediction of solar and wind energies and load demand based on ML, such as ANN and support vector machine (SVM), can be combined with the conventional methods for optimal energy management [24]. Among these methods, FLC, ANN, and ANFIS have been popular in recent years. Table 1 shows the advantages and disadvantages of these three methods, also compared to the RL-based method. The intelligent control strategies are able to manage the dynamic behavior of the hybrid system without exact mathematical models. However, these methods are not able to guarantee the optimal performance of the HRES [24].

Table 1. The advantages and disadvantages of some recently developed methods.

Methods	Advantages	Disadvantages
fuzzy logic control (FLC)	-Following the rule basis and membership functions (MF), easy to understand. -Insensitive to variation of the parameters. -Do not need a good model of the system and training process.	-Trial-and-error method for determining the MFs, time-consuming and not optimal performance. -Greater number of variables makes it more complex to optimize the MFs.
ANN	-Able to learn and to process parallel data. -Nonlinear and adaptive structure. -Generalization skills and design do not depend on system parameters. -Fast response capacities compared to the conventional method.	-Its "black box" nature and the network instruction problem lead to a lack of rules for determining the structure (cell and layers). -Historical data provides a need for the learning and tuning process. -The number of data set used to train the ANN defines the optimality.
adaptive neuro-fuzzy inference system (ANFIS)	-Has the inference ability of FLC and able to learn and process parallel data as ANN. -Applies neural learning rules to define and tune the MF of the fuzzy logic.	-More input variables lead to a more complex structure.
reinforcement learning (RL)	-Conducts learning without prior knowledge. -Can be combined with ANN for deep RL to solve the continuous state-space control problems.	-Long-time convergence for large real-world problem if not good initialization.

With technological development, ML has recently been applied in various areas. Researchers have been gradually shifting their interest towards studying the agent-based learning machine method for hybrid energy management, especially for the state-of-art RL and deep reinforcement learning (DRL) [25,26]. This subsection focuses on the summary of EMS based on RL.

Reinforcement learning is a heuristic learning method that has been applied to various areas [12]. The general model of RL is shown in Figure 1, which consists of the agent, environment, actions, states, and rewards. The purpose of RL is for the agent to maximize the reward by continuously taking actions in response to an environment. The next action can be defined based on the rewards and exploration-exploitation strategies like ε-greedy or softmax [16]. Q-learning is one of the most popular model-free RL algorithms. DRL is the combination of RL and the perception of deep learning. DRL has successfully performed in playing Atari and Go games [27]. In addition, DRL is a powerful method used to handle complex control problems and large state spaces by using a deep neural network to

calculate the value estimation and associated the pairs of state and action. Thus, the DRL method has been rapidly applied in robotics [27], building HVAC control [28], hybrid electric vehicles [29], etc.
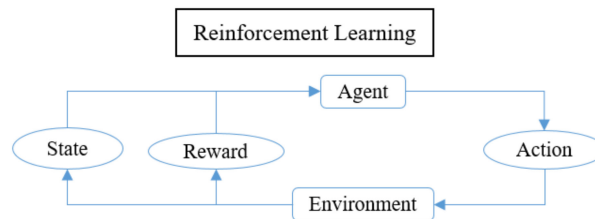

Figure 1. Scheme of reinforcement learning.

Some researchers have studied the use of RL and DRL energy management systems for hybrid electric vehicles and smart building [30,31]. However, few publications study on the energy management of the HRES. Kuznetsova (2013) proposed a two step ahead Q-learning method for defining the battery scheduling in a wind system, while Leo, Milton, and Sibi (2014) [32] developed a-three-step-ahead Q-learning for controlling the battery in a solar system. A novel online energy management technique using RL was developed in reference [33], which can learn and give the minimum power consumption without prior information on the workload. Additionally, a single agent system based on Q-learning has been developed by Kofinas (2016) for energy management of a solar system [34]. Finally, a fuzzy reward function has been introduced based on the Q-learning algorithm by Kofinas (2017) [35] to enhance the learning efficiency for controlling the power flow between components including PV, a battery, the local consumer, and a desalination unit for water supply.

A multi-agent system (MAS) includes a set of agents which interact with each other and with their environment. Due to its feature of solving complex problems in a more computationally efficient manner compared to a single-agent system, many researchers have used it to solve energy management problems [36]. A MAS-based system was considered in a grid-connected microgrid for optimal operation [37]. Additionally, a MAS-based intelligent EMS for the islanded microgrid is designed in reference [38] to balance the energy among the generators, batteries and loads. an autonomous multi-agent system for optimally managing the buying and selling power has been proposed by Kim (2012) [39]. Foo, Gooi, and Chen (2014) [40] introduced a multi-agent system for an energy generation and energy demand schedule. Following the EMS based multi-agent, a similar concept—energy body (EB)—was developed, in which the EB acts as an energy unit that has many functionalities and plays multiple roles at the same time [41,42]. The energy management problem (EMP) of energy internet (EI) has been defined as a distributed nonlinear coupling optimization problem in reference [42] and solved by the alternating direction method of multipliers algorithm. Moreover, the problem of day-ahead and real-time cooperative energy management has been successfully solved by the event-triggered-based distributed algorithm for the multi-energy system, formed by various EBs [41]. Multi-agent based energy management has been considered to be a potential and optimal solution to the control problem for microgrids. As shown in the literature review, most of the works based on the MAS approach tried to develop the mathematical models of the systems and solve the optimization problems. Taking the benefits of reinforcement learning into account, some authors have proposed the MAS approach with learning abilities which can reduce the task of system modeling and complex optimization problems. A multi-agent system using Q-learning has been developed by Raju (2015) [43] to reduce the solar system's energy consumption from the grid. Finally, Kofinas (2018) [44] has been proposed a cooperative multi-agent system based on Fuzzy Q-learning for energy management of a standalone microgrid.

To overcome the disadvantages of the Q-learning method in practical applications which can only handle the discrete control problems, a deep Q-learning algorithm is introduced to reduce the problem with large state-action pair. In Q-learning, Q-values are saved and updated for each state-action pair. However, in deep Q-learning, the neural network is used in the good approximations in the Q-function

for the continuous state-space problems. The model is a convolutional neural network, which is trained with a variant of Q-learning. The framework of deep Q-learning is shown in Figure 2. A deep neural network, which can estimate the state of environment in the next step, is used to improve the convergence rate of the Q-learning. Based on Bellman's equation, we can calculate the loss function by taking the mean-square error (MSE) between the Q-value estimated by neural network and the result from the Bellman's equation. Figure 3 shows the hybrid renewable hydrogen energy system while Figure 4 is the conceptual scheme of the power management control based on the deep Q-learning method. The system will be developed in this project for the improvement of a power system in Basco Island.
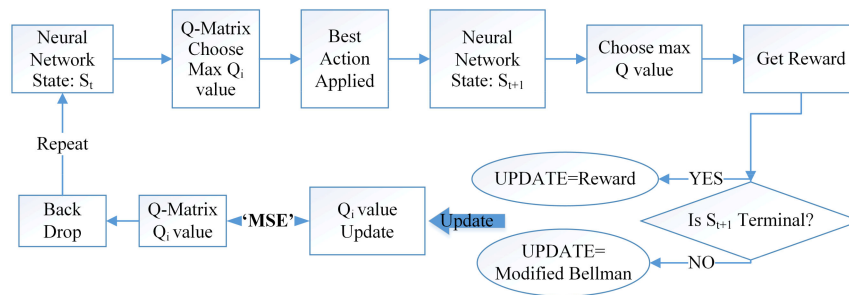

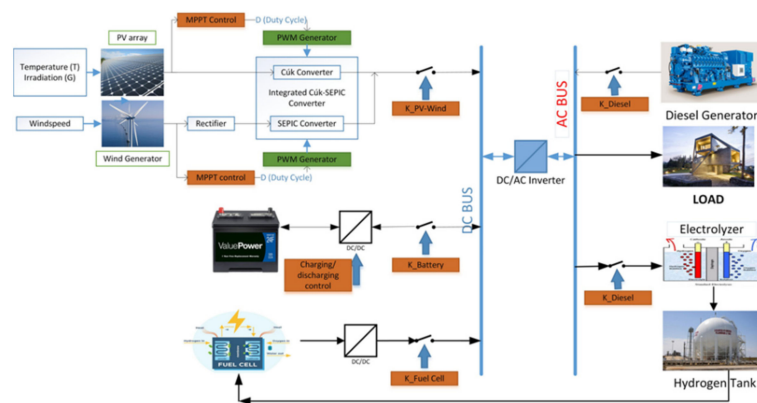**Figure 2.** The framework of deep Q-learning.
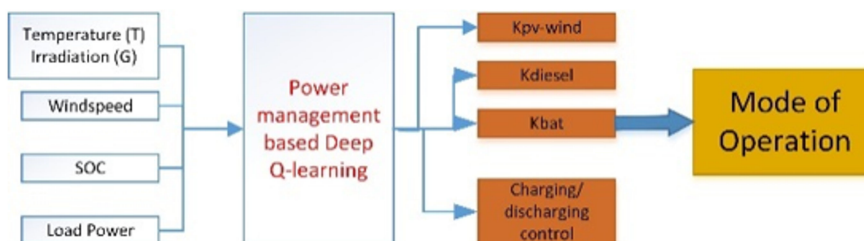

Figure 3. The proposed HRES.

Figure 4. EMS based on Deep Q-learning of a hybrid renewable energy system (HRES).

3. Optimal Sizing of HRES Based on HOMER

3.1. Site Description

In this section, a feasible study of HRES was carried out by HOMER to improve the isolated microgrid for cost efficiency and sustainable development. Detailed steps of the system design for optimal configuration using HOMER are illustrated in Figure 5 [45]. The selected location is Basco island, located in the northern region of the Philippines about 190 km away from Taiwan. Farming and fishing are the two major economic sectors in this area. Currently, the island is powered by a diesel

generator system with high operational costs. Figure 6 shows the fuel supply chain in this island. Due to the excellent location of the island for marine resource management and tourism, the demand for the sustainable economic development forces the local government to develop a new reliable and environmentally-friendly system for power supply to the local community. Figure 7 indicates the schematic of the proposed energy system and the actual yearly load profile of the Basco island, while Figure 8 illustrates the typical daily load profile with an average demand of about 700 kWh. Following the data, the power system must supply about 18 MWh per day with a peak of about 1.4 MW. To fulfill the load demand in this area, a new HRES is proposed including solar and wind generators, diesel generator, hydrogen system, and batteries. As shown in Figure 7, the system consists of a 220V AC bus and a 48V DC bus. To exchange the power, a bidirectional inverter is installed between the AC bus and the DC bus.
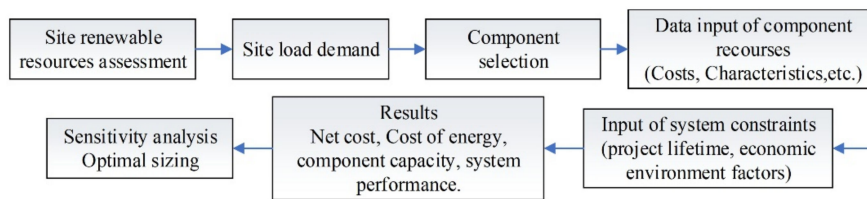

Figure 5. Detailed steps of HOMER used for studying optimal sizing analysis.

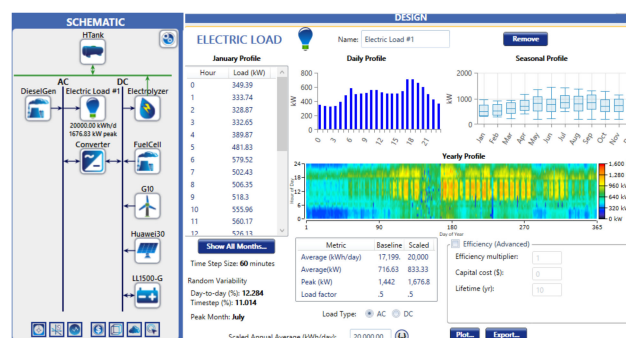**Figure 6.** Fuel supply chain in Basco.

Figure 7. The schematic of the proposed HRES and input of Basco load demand.
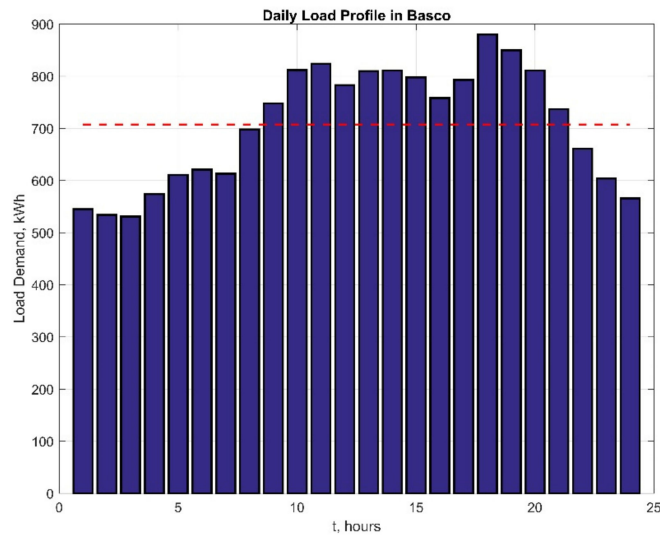
Figure 8. The typical daily load profile in Basco, Philippines.

In this project, weather data were taken from the National Renewable Energy Lab database (NREL) for system simulation. As indicated in Table 2, the average solar radiation every year is around 4.44 kWh/m^2/day while that of wind speed is 7.22 m/s.

Table 2. Weather data in Basco Island.

Month	Daily Solar Radiation (kWh/m^2/day)	Ambient Temperature (°C)	Average Wind Speed (m/s)
January	3.149	23.40	9.33
February	3.739	23.41	8.39
March	4.834	24.17	6.88
April	5.262	25.29	5.86
May	5.939	26.54	4.95
June	5.229	27.03	5.57
July	5.378	27.11	5.58
August	4.966	27.17	5.60
September	4.529	27.24	6.05
October	4.079	27.11	8.47
November	3.194	26.00	9.96
December	2.993	24.34	10.04

3.2. System Components

The cost and characteristics of each component, such as lifetime, efficiency, and power curve, need to be figured out for the calculation in HOMER. Table 3 shows all the kinds of components used in the project, including their technical specifications, economic costs (investment cost, replacement cost, operation and maintenance cost), and the search spaces of their capacity.

Table 3. Technical and economic specifications of the system components.

PV	Generic Flat Plate PV
Factors	**Value**
Nominal power	1 kW
Materials	Polycrystalline silicon
Derating factor	80%
Slope	21 degree
Ground reflection	20%
Lifetime	25 years
Capital cost	2500 US$/KW
Replacement cost	2250 US$/KW
Operation and Maintenance (O&M) cost	10 US$/kW/year
Search space	0~15,000 kW
Battery	**Generic 1 kWh Lead Acid**
Factors	**Value**
Nominal capacity	1 kWh
Maximum capacity	83.4 Ah
Nominal voltage	12 V
Maximum charge current	16.7 A
Maximum discharge current	24.3 A
Maximum charge rate	1 A/Ah
Lifetime	8 years
Capital cost	700 US$/unit
Replacement cost	500 US$/unit
O&M cost	10 US$/year
Search space	0~25,000 kW
Electrolyzer	**Generic**
Factors	**Value**
Lifetime	25 years
Capital cost	2250 US$/kW
Replacement cost	2025 US$/kW
O&M cost	0.1 US$/op. hr.
Search space	0~5000 kW
Hydro tank	
Factors	**Value**
Lifetime	25 years
Capital cost	2250 US$/kW
Replacement cost	2025 US$/kW
O&M cost	0.1 US$/op. hr.
Search space	0~5000 kW
Wind Turbines	**Generic 10 kW**
Factors	**Value**
Rotor diameter	3 m
Rated power	10 kW DC (at 12.5 m/s)
Voltage	48V DC
Lifetime	25 years
Starting wind speed	3.31 m/s
Cut-off wind speed	15 m/s
Capital cost	50,000 US$/unit
Replacement cost	45,000 US$/unit
O&M cost	500 US$/year
Search space	0~1000 units

<div align="center">**Table 3.** *Cont.*</div>

Diesel Generator	Generic Large Genset
Factors	**Value**
Minimum load ratio	30%
Lifetime	15,000 h
Fuel	Diesel
Capital cost	1000 US$/kW
Replacement cost	750 US$/kW
O&M cost	0.5 US$/op. hr.
Search space	0~750 kW
Fuel Cell	**Generic fuel cell**
Factors	**Value**
Minimum load ratio	25%
Lifetime	40,000 h
Fuel	Hydrogen
Capital cost	2,250 US$/kW
Replacement cost	2,025 US$/kW
O&M cost	0.1 US$/op. hr.
Search space	0~5000 kW
Converter	**Generic**
Factors	**Value**
Lifetime	25 years
Efficiency	95%
Capital cost	1000 US$/kW
Replacement cost	9000 US$/kW
O&M cost	0 US$/year
Search space	0~5000 kW

3.3. Optimization Criteria

The criteria for choosing the optimal sizing of the hybrid renewable power system are usually influenced by the economic and power reliability factors. Generally, according to this method, we can find out the suitable combination of system components and their capacity, including the lowest net present cost (NPC) and cost of energy (COE), which can meet the load demand at all times.

3.3.1. The Net Present Cost

The NPC is considered as the sum of all the relating cost in the project lifetime and is computed by the following equations [46]:

$$NPC = \sum_{N=1}^{N=t} f_{d,N} \left(C_{cap} + C_{rep} + C_{main} - C_s \right) \tag{1}$$

where t is the project lifetime. C_{cap}, C_{rep}, C_{main}, C_s are the capital, replacement, Operation and Maintenance (O&M), and salvage cost, respectively. $f_{d,N}$ is calculated by [46]:

$$f_{d,N} = \frac{1}{(1+i)^N} \tag{2}$$

where i and N are the annual interest rate and the year when the calculation is performed, respectively.

3.3.2. Cost of Energy

The COE in HOMER is defined as the average cost per kWh of served electric energy E_{served} and is determined by [46]:

$$COE = \frac{AC_T}{E_{served}} = \frac{\sum C_{acap} + C_{arep} + C_{amain} - C_{as}}{E_{served}} \tag{3}$$

where AC_T is the total cost of the component "a" of the project lifetime at each year, and C_{acap}, C_{arep}, C_{amain} and C_{as} are the related cost of component "a".

3.4. Optimal Sizing Results

Following the weather data and load profile collected from the site, the project lifetime of this study was considered with the value of 25 years while the discount rate and inflation rate are 7.5% and 3%, respectively. The constraint of the minimum renewable fraction of the system was set to 70%.

According to the calculation results, the optimal configuration is defined among all the feasible configurations, in which the values of NPC and COE of the system are about 72.5 million US$ and 0.696 US$/kW, respectively. Additionally, the operation cost of the system is more than 1.9 million US$. The optimal configuration of the proposed system for the case study at Basco Island includes 5483 kW of PV, 236 units of 10 kW Wind turbines, 20,948 kW of batteries (48V DC, 4 modules, 5237 strings), 500 kW of Fuel Cells, a 750 kW Diesel generator, a 3000 kW Electrolyzer, a 500 kg H-tank, and a 1575 kW Converter. The total electric production is about 13.8 GW/year and the excess energy is around 11.2%. As can be seen from Figure 9, the monthly average electric production is illustrated. WT produces more energy in winter and spring, while solar PV generates more power in summer and autumn.
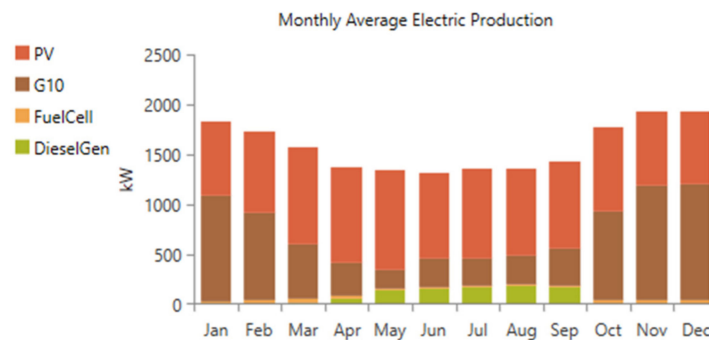

Figure 9. Monthly average electric production.

It is indicated in Table 4 that the percentages of power production of the primary resources are 54.4% and 39.3% for PV and a wind turbine, respectively. Based on the hydrogen production as shown in Figure 10, the contribution of the fuel cell is 1.58% of total production. With the support from PV and WT as the primary power generators and fuel cells and batteries as the storage system, the use of diesel generator reduces by about 4.8%. It can be concluded that this is a high renewable fraction power system, which provided a for 91% RE. Thus, the amount of greenhouse gas emissions can be significantly decreased as shown in Table 5, compared to the case of a full diesel generator being used. Figure 11 illustrates the cash summary of all components in the optimal configuration, including capital, replacement, O&M, fuel, and salvage costs. It can be seen from Table 6 that most of the total NPC is for PV and wind turbines, accounted for 18.8% and 17.5%, respectively, due to their high investment cost. However, the highest contribution of NPC belongs to the battery with a value of around 41% because of its short life over the 25-year project. It is clear that diesel generator also has high NPC with a value of about 11% despite its low investment cost. This is because of the high cost of fuel with a value of more than US$2.4 million.

Table 4. Electrical production.

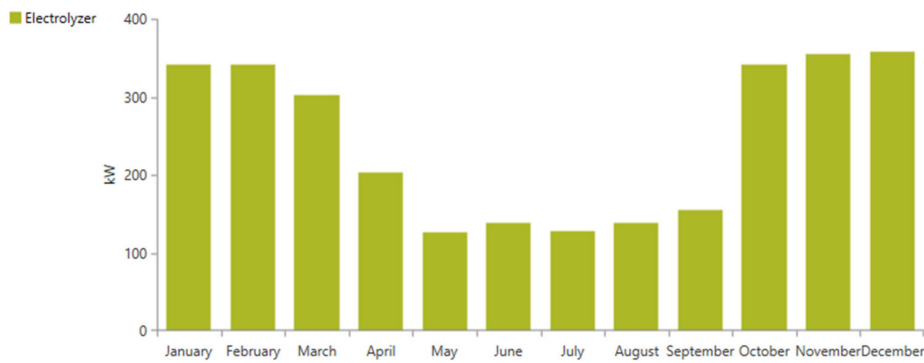| Component	Production kWh/year	Percentage (%)
Generic flat plate PV	7,510,627	54.4
Generic 10 kW WT	5,421,873	39.3
Diesel generator	660,222	4.7
Fuel cell	218,542	1.6
Total	13,811,263	100

Figure 10. The hydrogen production of the optimal configuration.

Table 5. The emissions of the optimal configuration.

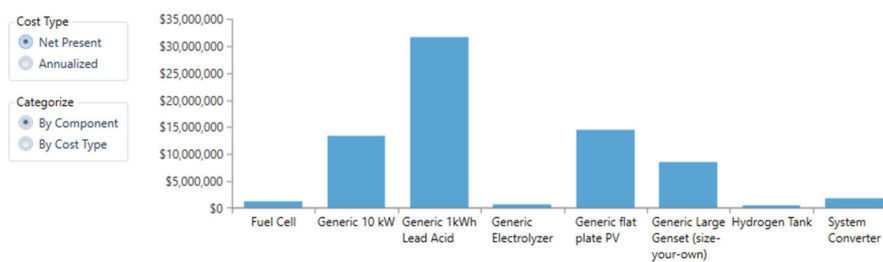| Emission Factors	Proposed HRES	100% Diesel Generator	Units
Carbon dioxide	448,527	5,098,748	kg/yr.
Carbon monoxide	2320	26,378	kg/yr.
Unburned hydrocarbons	123	1400	kg/yr.
Particulate matter	19.8	226	kg/yr.
Sulfur dioxide	1096	12,464	kg/yr.
Nitrogen oxides	445	5056	kg/yr.

Figure 11. The cost summary of system components.

Table 6. Detailed costs of system components.

Component	Capital (US$)	Replacement (US$)	O&M (US$)	Fuel (US$)	Salvage (US$)	Total (US$)
PV	13,707,782	0	782,321	0	0	14,490,104
WT	11,800,000	0	1,683,604	0	0	13,483,604
DG	750,000	317,698	5,072,215	2,440,485	−86,186	8,494,213
Battery	14,663,600	16,041,754	2,988,826	0	−2,122,299	31,571,880
Fuel cell	1,125,000	0	415,194	0	−195,842	1,344,351
Electrolyzer	750,000	0	0	0	0	750,000
H_2 tank	500,000	0	0	0	0	500,000
Converter	1,757,266	546,895	0	0	−323,253	1,798,909
System	44,871,649	16,906,349	10,942,162	2,440,485	−2,727,582	72,433,063

4. The Proposed h-POQL MPPT Control

4.1. The Assessment of the MPPT Control Methods

The power generated by the PV and wind turbine systems is strongly dependent on the weather conditions. Thus, the hybrid system requires power converters to change the power forms and to transfer efficiently by applying MPPT techniques to extract maximum energy from wind and solar. The following process is the concept of MPPT control.

- In Figure 12a, based on a typical solar radiation and temperature, there is a unique maximum power point (MPP) on the power-voltage (P-V) curve where the system can operate at the maximum efficiency and produce maximum power. Similar to PV system, the wind turbine produces maximum output power at a specific point of P-ω_m curve as shown on the right hand side of Figure 12b. Thus, it is necessary to continuously track the MPP in order to maximize the output power. In generally, the major tasks of MPPT controller include:

    1. How to quickly find the MPP.
    2. How to stably stay at the MPP.
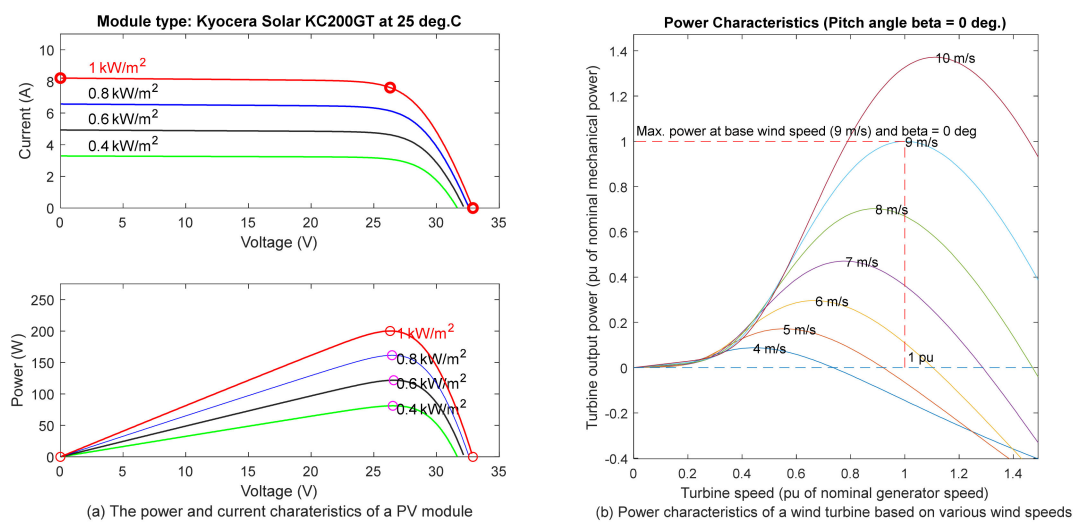   3. How to smoothly move from one MPP to another for rapid weather condition change.

Figure 12. The output power of the solar panel at 25 °C with different solar radiation and the output power of the wind turbine (WT) with various wind speeds.

Based on numerous studies of MPPT in the last few decades, the comparison between these approaches are shown as follows [14,15]:

- Conventional methods, such as Perturbation & Observation (P&O), Incremental Conductance (IC), Open Circuit Voltage (OV), and Short Circuit Current (SC), are famous for their easy implementation, but their disadvantages are that they are poor convergence, slow tracking speed, and high steady-state oscillations. In contracts, AI methods are complicated in design and require high computing power. However, due to the technological development of computer science, the AI method-based MPPT methods are a new trend with fast tracking speed and convergence, and low oscillation [15].
- A lot of MPPT methods have been developed following soft computing techniques, including FLC, ANN, and ANFIS [47]. The drawbacks of these methods are that they need a large computer memory for training and the rule implementation.

- The next era of MPPT control is based on the evolution algorithms such as Genetic Algorithm, Cuckoo Search, Ant Colony Optimization, Bee Colony, Firefly Algorithms, and Random Search since these methods can efficiently solve the non-linear problems. Among these methods, PSO has become more commonly used in this field due to its easy implementation, simplicity, and robustness. Besides, it can combine with other methods to create new approaches [15,47].
- Hybrid methods which integrate two or more MPPT algorithms together have a better performance and utilize the advantages of each method such as PSO-P&O, and PSO-GA [15]. The advantage of these methods is that they can help to track the global maximum power point quickly under the partial shading conditions.

To overcome the disadvantages of these recent MPP methods, some researchers have focused on the field of Q-learning to handle the MPPT control problems. In reference [48], Wei has developed Q-learning algorithm for MPPT control of variable-speed WT system, and Youssef applied the method for online MPPT control [17]. In addition, some researchers from National Chiayi University in Taiwan have proposed a RL-based MPPT method for the PV system [16]. One of the latest research examples in this area can be found in reference [18] where the authors proposed a new Q-learning based MPPT method for the PV system with larger state spaces, compared to only four states in reference [17] and reference [16]. The simulated results with good system performance from these papers show that the application of RL in the field of MPPT control is emerging and promising, which can help to improve the efficiency in renewable energy conversion, especially for solar and wind energy systems.

Q-learning is a useful RL method for handling and figuring out the running average values of the reward function. Considering that S is a discrete set of states, where A is a discrete set of actions, the agent will experience every state $s \in S$ and possible set of actions $a \in A$ through the learning process. When taking the action a_t the agent will transit from state s_t to state s_{t+1} and receive a reward r_{t+1}, then the Q-learning update rule is given by equation below [48]:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha\left[r_{t+1} + \gamma \max_{a_i} Q_t[s_{t+1}, a_i] - Q_t[s_t, a_t]\right] \tag{4}$$

in which, $Q_t(s_t, a_t)$ is the action value function, α is the learning rate, γ is the discount factor, and $\max_{a_i} Q_t(s_{t+1}, a_i)$ is the maximum expected future reward given the new state s and possible action at the next step. The flowchart of Q-learning algorithm is demonstrated in Figure 13 [12].

The output power of PV system can be calculated by the equation below [16]:

$$P_{pv} = V_{pv}I_{pv} = V_{pv}\left\{I_{ph} - I_{pvo}\left[e^{\frac{q}{AkT}[V_{pv}+I_{pv}R_s]} - 1\right]\right\} \tag{5}$$

where I_{ph} is the light-generated current, R_s is the series resistance, A is the non-ideality factor, k is the Boltzmann constant, I_{pvo} is the dark saturation current, T is temperature, and q is the electron charge.

Generally, there are two stages in the MPPT control based on Q-learning: the offline learning process and the online application process [12]. Firstly, the agent will learn a map from state to action and then the learned values of the actions will be stored in the Q-table. Following this Q-table, the relationship between the voltage and power is determined. Secondly, the action value Q-table will be used to control the PV system in the application process. The procedure of initial input configuration for Q-learning is shown in Figure 13 as follows [16]:

- State-spaces are represented by the voltage-power pair:

$$S = \left\{s \big| s_{kj} = (V_{pv,k}, P_{pv,j}), k \in [1, 2, \ldots, N], j \in [1, 2, \ldots, M]\right\} \tag{6}$$

- Action-spaces are the perturbations of duty cycle ΔD to the PV voltage:

$$A = \{a | {+}\Delta D, 0, {-}\Delta D\} \tag{7}$$

- Rewards:

$$r_{t+1} = \begin{cases} w_p \Delta P \ \ if \ \Delta P > \delta_1 \\ w_{best} \ \ if \ |\Delta P| > \delta_1 \ and \ a_i \neq 0 \\ w_n \Delta P \ \ if \ \Delta P < \delta_1 \ or \ a_i = 0 \end{cases} \tag{8}$$

where $\Delta P = P_{t+1} - P_t$ and δ_1 is the small number represented as the small area around the maximum power point. Based on the weights w_p, w_{best}, and w_n the separation between positive, best, and negative states is clearly defined.
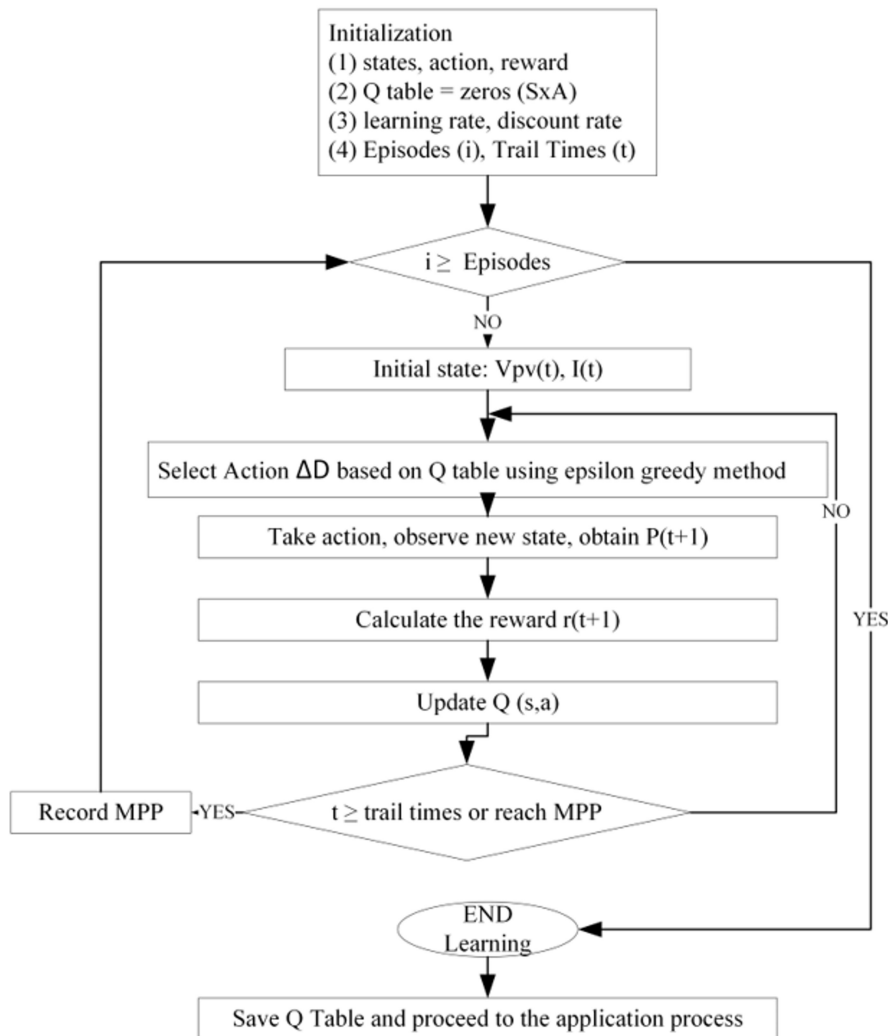

Figure 13. The learning state of the maximum power point tracking (MPPT) based Q-learning algorithm.

Based on the state of art reinforcement learning in the field of MPPT control, the proposed h-POQL method aims to get the advantages of low learning time, low cost, and easy implementation in a practical system. By separating the control regions based on the irradiation and temperature, the state space can be reduced. The agent will spend less time learning the optimal policy in a small control region. In addition, the fixed step size of the duty cycle is the major problem of the P&O method in the response to the fast change of weather conditions. The Q-learning method aims to use a variable step size to define the optimal duty cycle in a specific control region. With the knowledge learned by the Q-learning agent, the P&O can change the reference input of the duty cycle so that the smaller step size of the duty cycle can be applied to track for the maximum power of the PV source.

4.2. Methodology of the h-POQL MPPT Control

Following the previous review on the MPPT methods, this work proposes a simple hybrid MPPT control method, which is the combination of Q-learning and P&O, to overcome the disadvantages of each technique. In MPPT based on the P&O as shown in Figure 14, the oscillation with large step perturbation around the maximum power point and the low response to the change of weather conditions are the main constraints. On the other hand, the method following the Q-learning algorithm can just handle the discrete states and actions, so longer computational time in case of large states spates is the major limitation. Details of the h-POQL method will be described below.
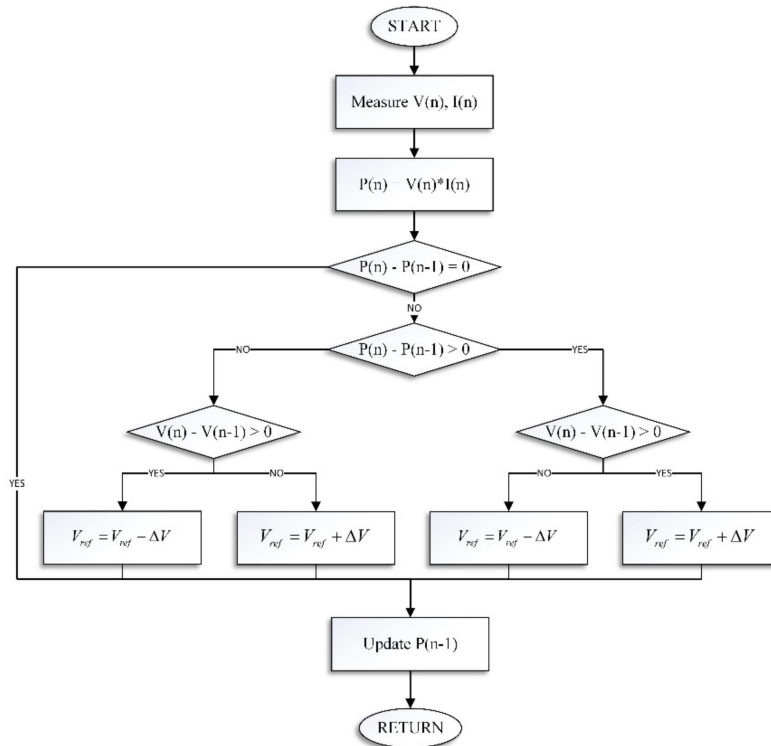
Figure 14. Flow chart of the perturb and observe algorithm.

The proposed h-PORL MPPT method can be shown in Figure 15. As shown in Figure 16, it can be divided into eight control zones based on the temperature and irradiation. In each control zone, the Q-leaning-based MPPT method will learn the responses of the PV source for the optimal values of the duty cycle. Then these optimal values will be used as the inputs for the P&O MPPT controller. This study aims to reduce the learning time by decreasing the number of discrete state spaces, and to improve the P&O MPPT method by lowering the variable step size. As shown in Figure 17, the testing model built in Simulink is the combination of the Kyocera solar KT200GT module, one boost converter, and one resistor which acted as the load.
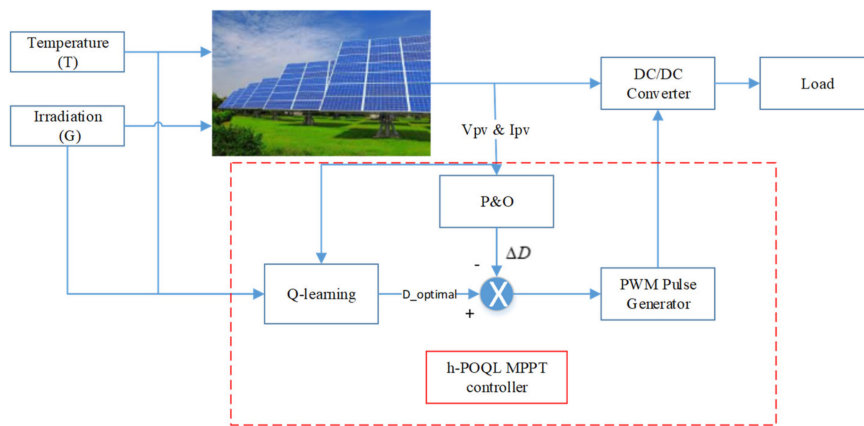
Figure 15. Block diagram of the proposed h-POQL MPPT method.

		Solar Irradiation (G, W/m²)			
		300–500	500–700	700–900	900–1100
Temperature (°C)	10–30	QL_1	QL_2	QL_3	QL_4
	30–50	QL_5	QL_6	QL_7	QL_8

Figure 16. Control zones of the PV system based on the Q–learning algorithm.
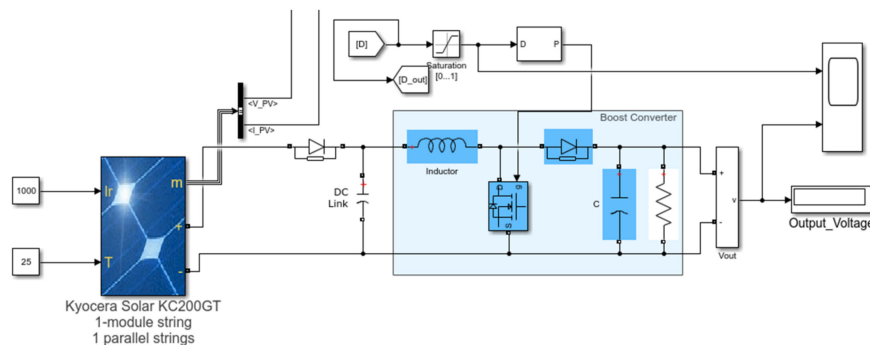
Figure 17. The PV system model in Simulink.

4.3. Simulation Results

4.3.1. Simulation of MPPT Control Based on Q-Learning

First of all, the Q-learning MPPT controller will be simulated and tested based on the data from the standard testing conditions (STCs), which are 1000 W/m² irradiation and the 25°C panel temperature. In each episode, the maximum training time is set to 5 s and stops when the maximum power point is reached. The whole training process will finish when all the episodes are conducted. Figure 18 indicates the good performance of the controller. Due to the update of the Q-table, the training process tends to reduce over the training period. Following the duty cycle value of 39.5%, the output power of the PV module is around 200.2W, which is almost equal to the data from the manufacturer with the value of 200.14 W.
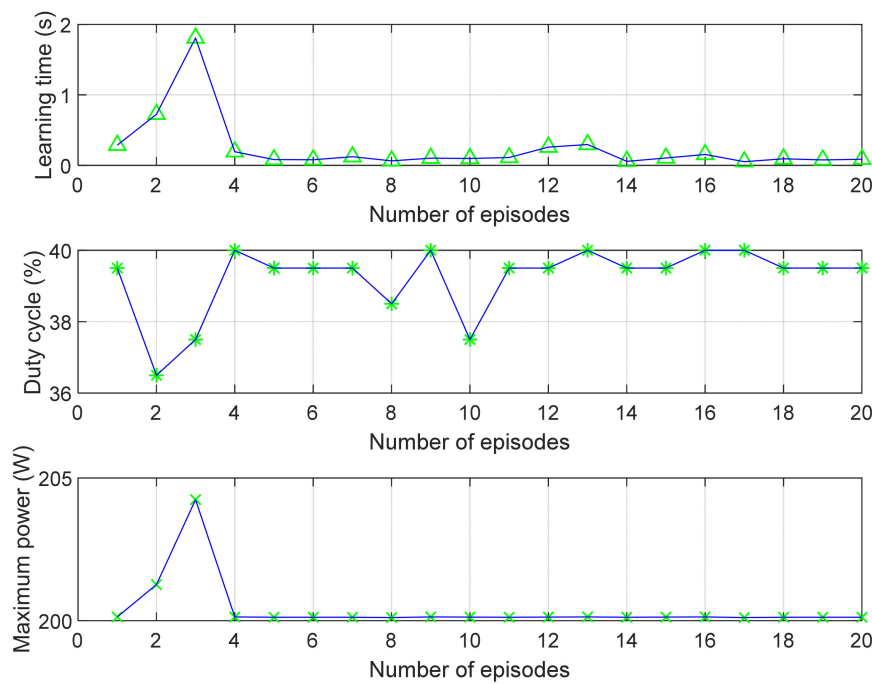
Figure 18. Q-learning MPPT training based on the standard test conditions (G = 1000 W/m^2 and T = 25°C).

4.3.2. Simulation and Validation of h-POQL MPPT Controller

In this section, eight Q-learning controllers in the relative control zones were trained to find the optimal values of the duty cycle. The simulated results are shown in Table 7. In the next stage, different operating conditions are used to evaluate the performance of the h-POQL controller. First, the temperature of the power source is set to 25 °C, and the irradiation is switched between 450, 650, 750, and 950 W/m^2. Later, the irradiation is fixed to 1000 W/m^2 and the temperature changes between 15 °C and 35 °C. Results in Figures 19 and 20 show that for all cases the controller can perform with the fast convergences to the steady state and operates at the maximum power point condition, compared to the theoretical data of the PV module.
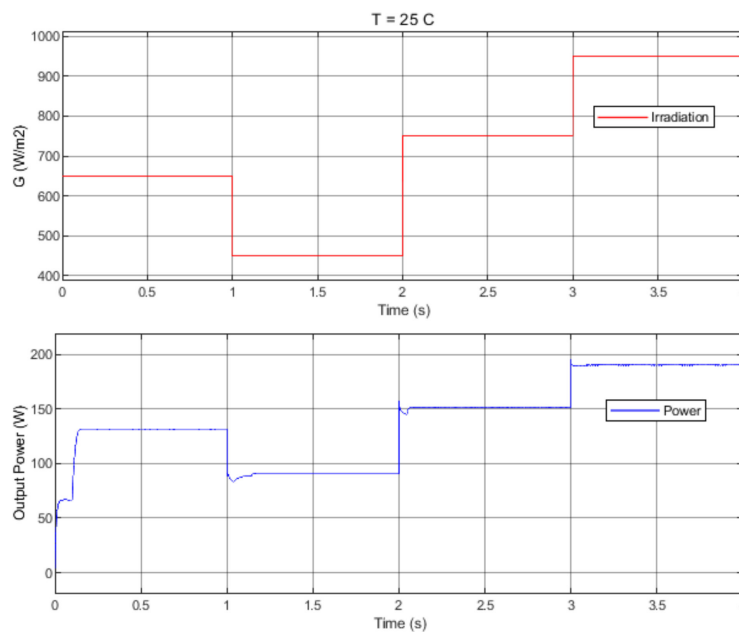

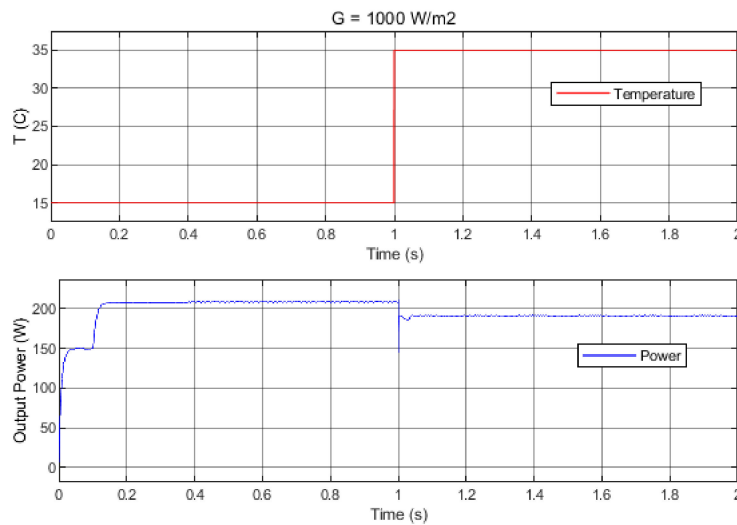**Figure 19.** Output powers under the change of irradiation.

Figure 20. Output powers under the change of temperature.

Table 7. Optimal duty cycle in eight control zones.

	QL1	QL2	QL3	QL4	QL5	QL6	QL7	QL8
Duty cycle (%)	17	21	32	39	19	24	35	41

Finally, the proposed hybrid controller is compared with the P&O method based on the change of both temperature and irradiation, as shown in Figure 21. The results in Figure 22 illustrates that the step size of the P&O can be reduced from 0.0005 to 0.00005 in the h-POQL controller. Thus, it can overcome the oscillation drawback of the P&O method. Moreover, more power was generated by h-POQL controller with the change of weather conditions as indicated by the blue line in the graph. In conclusion, a better performance of the h-POQL over the P&O can be validated.
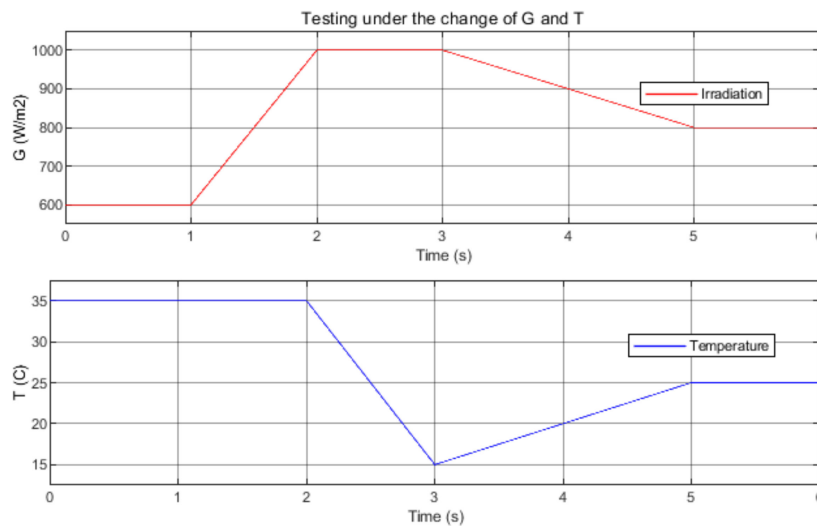

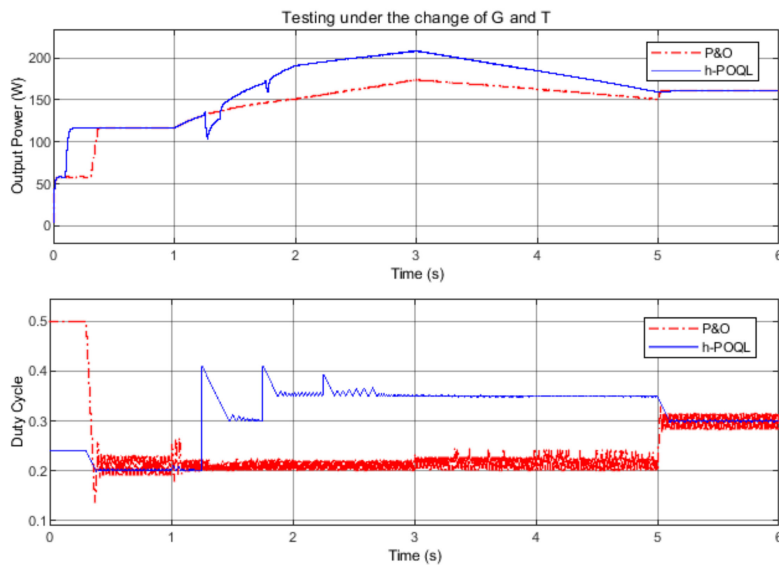**Figure 21.** The change of weather conditions.

Figure 22. Comparison between the h-POQL and the P&O methods.

5. Discussions

This paper provides the assessment of hybrid renewable hydrogen energy system development, especially for the practical application of rural and islanded electrification. Most remote areas are currently powered by diesel generators that could significantly pollute our environment. With the development of new technologies, the cost of renewable energy will probably decrease allowing HRESs to be implemented for sustainable development. Optimal sizing of the system helps to define the optimal configuration that can ensure the power supply with the lowest cost, while the MPPT control and EMS are essential to maximize the harvested power and to control the power flow among the various components in the system. Based on the successful applications of reinforcement learning in various fields, the system could be a possible solution to the problems involved in the hybrid renewable energy system design.

In recent times, various methodologies have been applied to size the system components so as to minimize the cost, ensure the reliability and reducing the emissions, the HOMER methodology is one of the most popular of these methodologies. A detailed process for optimally using HOMER was clearly indicated with the case study in Basco Island. As mentioned above, the major drawbacks of battery are a short lifetime and recycling problems, so the focus on the development of hydrogen systems combined with renewable resources should be significantly considered as alternatives to fossil fuel and nuclear power. Moreover, the analytical techniques or tools are necessary for solving the optimization problem in system sizing based on the design criterions and constraints. Huge research has been carried out based on various tools and techniques. AI techniques are able to completely search the workspace and to define the global optimal solution, but sometimes they also inefficiently solve certain difficulties when increasing the number of variables. For overcoming the limitation of sizing problems, ML and RL techniques, as well as the hybrid methodology, should be focused on.

The main objectives of an MPPT controller are to deal with the problems involved in the fluctuation and intermittency of RE sources due to the change of weather conditions while EMS is used to optimize operations, ensure the system reliability, and provide power flow control in both standalone and grid-connected microgrids. In this study, the proposed h-POQL method was developed for MPPT control of the PV source. Based on the simulated results, the proposed method can efficiently track the maximum power under various changes to the weather conditions. In addition, it shows better results in terms on speed and accuracy when the h-POQL is compared against the P&O method. The Q-learning controllers have been trained offline for different desired targets, such temperature and irradiation, and then we transferred the training models to the P&O controller to increase the efficiency

of energy conversion. In contrast, the approach in reference [18] adopted Q-learning as an on-policy algorithm. Due to the different approaches between two studies, the comparison with the method in reference [18] was not carried out. However, based on the simulation results, the proposed h-POQL has clearly shown faster response based on the change of weather conditions, with less than a second compared to more than two seconds [18], meaning h-POQL could be more efficient. This is because the controller in the previous paper needs to spend time on the online learning. In future work, the real experiment will be set up for testing the h-POQL algorithm, and the comparison between these two methods will be conducted.

Following the assessments of EMS and MPPT control conducted in this study, there is a trend in the application of ML and RL algorithms in this field. Most of the current work just focuses on the simulation, so real-time experiments should be implemented to verify the performance of the agent-based learning techniques for the improvement of energy conversion and management. With the feature of a self-learning ability, multi-agent-based energy management based on RL was proven to have potential and be effective in supervisory and local control, but there is still a need to improve the communication mechanism between agents in the control system. Finally, it has been shown that the RL algorithm is has high performance, however, the discrete state spaces and actions are the major limitations of this method. Further study on the DRL for the control strategies of HRES should be explored to deal with the control problems of continuous state spaces and actions.

6. Conclusions and Future Works

This research aims to develop the hybrid renewable hydrogen energy system, especially for a standalone microgrid with the applications of rural and islanding electrification. The problems involved in the system design process were clearly introduced, including optimal sizing, MPPT control, and energy management system.

Firstly, according to the data collected from the Basco island in the Philippines, the optimal design of HRES was determined by the HOMER software which has the features of being cost-effective, reliable, and environmentally friendly. According to the analysis, the optimal configuration of power system includes 5483 kW of PV, 236 units of 10 kW of wind turbines, 20,948 kW of batteries (48V DC, 4 modules, 5237 strings), 500 kW of fuel cells, a 750 kW diesel generator, a 3000 kW electrolyzer, a 500 kg H-tank, and a 1575 kW converter with the energy cost of US$0.774/kWh based on a 1 US$/liter fuel cost. Moreover, from the analyzed results, the combination of the fuel cell system and the battery is one of the best options for the design of HRES, in which FC can be used as long term energy storage option and the battery can act as a short term energy storage medium. The system is not only practical and cost-effective but can also satisfy the load demand in the applied area. The same work can be considered for the other sites around the world, especially in remote areas, to efficiently increase the renewable energy use and reduce emissions.

In regard to the recent successful applications of RL techniques in various fields, especially the areas of computer vision and robotics, this research aims to consider these theories for the MPPT control and energy management of HRES. According to the brief review and comparison between techniques for MPPT control and EMS, from conventional methods to the current AI ones, this paper can be a good reference for researchers in this field. This work introduces a new hybrid approach for MPPT control based on the combination of Q-learning and P&O, named as h-POQL. The proposed method was simulated in Simulink with various scenarios based on the change of weather conditions to test its efficiency and performance. It also shows better results in terms of power generation and speed. Additionally, it can define the optimal duty cycle in a specific control region by reducing the redundant states. Based on the optimal results learned by the Q-learning algorithm, the P&O can tune the reference input values of the duty cycle and track the maximum power point with faster speed and higher accuracy.

Based on the ability to learn from experiences and optimally solving complex control problems with no prior knowledge of the environment or complex mathematical model needed, reinforcement

learning is supposed to be the new and potential trend in the fields of energy conversion and management. In the future, optimal sizing based reinforcement learning will be studied and compared with the approach from HOMER software in order to obtain optimal results and be able to meet more required variables and constraints. Then the practical system will be installed at the applied site when all the design requirements can be met. In addition, we plan to study more RL algorithms so that it can deal with continuous state-space problems besides the proposed h-POQL method. Further experiments will be implemented to test and compare the performance of these methods. Finally, the DRL algorithm will be integrated with the multi-agent-based HRES for energy management. Many real tests will be carried out for validation besides the simulation results. Our goal is to implement the proposed system on an isolated micro-grid.

References

1. ASEAN Center for Energy Team. *ASEAN Renewable Energy Policies*; ASEAN Centre for Energy: Jakarta, Indonesia, 2016.
2. Lin, C.E.; Phan, B.C. Optimal Hybrid Energy Solution for Island Micro-Grid. In Proceedings of the 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), Atlanta, GA, USA, 8–10 October 2016.
3. Chauhan, A.; Saini, R.P. A review on Integrated Renewable Energy System based power generation for stand-alone applications: Configurations, storage options, sizing methodologies and control. *Renew. Sustain. Energy Rev.* **2014**, *38*, 99–120. [CrossRef]
4. Vivas, F.J.; de las Heras, A.; Segura, F.; Andújar, J.M. A review of energy management strategies for renewable hybrid energy systems with hydrogen backup. *Renew. Sustain. Energy Rev.* **2018**, *82*, 126–155. [CrossRef]
5. Ahangari Hassas, M.; Pourhossein, K. Control and Management of Hybrid Renewable Energy Systems: Review and Comparison of Methods. *J. Oper. Autom. Power Eng.* **2017**, *5*, 131–138.
6. Fadaee, M.; Radzi, M.A.M. Multi-objective optimization of a stand-alone hybrid renewable energy system by using evolutionary algorithms: A review. *Renew. Sustain. Energy Rev.* **2012**, *16*, 3364–3369. [CrossRef]
7. Mellit, A.; Kalogirou, S.A.; Hontoria, L.; Shaari, S. Artificial intelligence techniques for sizing photovoltaic systems: A review. *Renew. Sustain. Energy Rev.* **2009**, *13*, 406–419. [CrossRef]
8. Siddaiah, R.; Saini, R.P. A review on planning, configurations, modeling and optimization techniques of hybrid renewable energy systems for off grid applications. *Renew. Sustain. Energy Rev.* **2016**, *58*, 376–396. [CrossRef]
9. Dawoud, S.M.; Lin, X.; Okba, M.I. Hybrid renewable microgrid optimization techniques: A review. *Renew. Sustain. Energy Rev.* **2018**, *82*, 2039–2052. [CrossRef]
10. Karami, N.; Moubayed, N.; Outbib, R. General review and classification of different MPPT Techniques. *Renew. Sustain. Energy Rev.* **2017**, *68*, 1–18. [CrossRef]
11. Olatomiwa, L.; Mekhilef, S.; Ismail, M.S.; Moghavvemi, M. Energy management strategies in hybrid renewable energy systems: A review. *Renew. Sustain. Energy Rev.* **2016**, *62*, 821–835. [CrossRef]
12. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2011.
13. Alpaydin, E. *Introduction to Machine Learning*; MIT Press: Cambridge, MA, USA, 2014.
14. Bendib, B.; Belmili, H.; Krim, F. A survey of the most used MPPT methods: Conventional and advanced algorithms applied for photovoltaic systems. *Renew. Sustain. Energy Rev.* **2015**, *45*, 637–648. [CrossRef]

15. Chandra, S.; Gaur, P.; Srishti. Maximum Power Point Tracking Approaches for Wind–Solar Hybrid Renewable Energy System—A Review. In *Advances in Energy and Power Systems*; Lecture Notes in Electrical Engineering; Springer: Singapore, 2018; Volume 508, pp. 3–12.

16. Hsu, R.C.; Liu, C.T.; Chen, W.Y.; Hsieh, H.I.; Wang, H.L. A Reinforcement Learning-Based Maximum Power Point Tracking Method for Photovoltaic Array. *Int. J. Photoenergy* **2015**, *2015*, 496401. [CrossRef]

17. Yousef, A.; El-Telbany, M.; Zekry, A. Reinforcement Learning for Online Maximum Power Point Tracking Control. *J. Clean Energy Technol.* **2015**, *4*, 245–248. [CrossRef]

18. Kofinas, P.; Doltsinis, S.; Dounis, A.I.; Vouros, G.A. A Reinforcement Learning Approach for MPPT Control Method of Photovoltaic Sources. *Renew. Energy* **2017**, *108*, 461–473. [CrossRef]

19. Indragandhi, V.; Subramaniyaswamy, V.; Logesh, R. Resources, configurations, and soft computing techniques for power management and control of PV/wind hybrid system. *Renew. Sustain. Energy Rev.* **2017**, *69*, 129–143.

20. Zia, M.F.; Elbouchikhi, E.; Benbouzid, M. Microgrids energy management systems: A critical review on methods, solutions, and prospects. *Appl. Energy* **2018**, *222*, 1033–1055. [CrossRef]

21. Jayalakshmi, N.S.; Gaonkar, D.N.; Nempu, P.B. Power Control of PV/Fuel Cell/Supercapacitor Hybrid System for Stand-Alone Applications. *Int. J. Renew. Energy Res.* **2016**, *6*, 672–679.

22. Roumila, Z.; Rekioua, D.; Rekioua, T. Energy management based fuzzy logic controller of hybrid system wind/photovoltaic/diesel with storage battery. *Int. J. Hydrogen Energy* **2017**, *42*, 19525–19535. [CrossRef]

23. Varghese, N.; Reji, P. Battery charge controller for hybrid stand alone system using adaptive neuro fuzzy inference system. In Proceedings of the 2016 International Conference on Energy Efficient Technologies for Sustainability (ICEETS), Nagercoil, India, 7–8 April 2016.

24. Chong, L.W.; Wong, Y.W.; Rajkumar, R.; Rajkumar, R.K.; Isa, D. Hybrid energy storage systems and control strategies for stand-alone renewable energy power systems. *Renew. Sustain. Energy Rev.* **2016**, *66*, 174–189. [CrossRef]

25. Buşoniu, L.; Babuška, R.; de Schutter, B. Multi-agent Reinforcement Learning: An Overview. In *Innovations in Multi-Agent Systems and Applications-1*; Srinivasan, D., Jain, L.C., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 183–221.

26. Nguyen, T.; Nguyen, N.D.; Nahavandi, S. Deep Reinforcement Learning for Multi-Agent Systems: A Review of Challenges, Solutions and Applications. *arXiv* **2018**, arXiv:1812.11794.

27. 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.

28. Mocanu, E.; Mocanu, D.C.; Nguyen, P.H.; Liotta, A.; Webber, M.E.; Gibescu, M.; Slootweg, J.G. On-line Building Energy Optimization using Deep Reinforcement Learning. *IEEE Trans. Smart Grid* **2019**, *10*, 3698–3708. [CrossRef]

29. Hu, Y.; Li, W.; Xu, K.; Zahid, T.; Qin, F.; Li, C. Energy Management Strategy for a Hybrid Electric Vehicle Based on Deep Reinforcement Learning. *Appl. Sci.* **2018**, *8*, 187. [CrossRef]

30. Fang, Y.; Song, C.; Xia, B.; Song, Q. An energy management strategy for hybrid electric bus based on reinforcement learning. In Proceedings of the 27th Chinese Control and Decision Conference (2015 CCDC), Qingdao, China, 23–25 May 2015.

31. Kim, S.; Lim, H. Reinforcement Learning Based Energy Management Algorithm for Smart Energy Buildings. *Energies* **2018**, *11*, 2010. [CrossRef]

32. Leo, R.; Milton, R.S.; Sibi, S. Reinforcement learning for optimal energy management of a solar microgrid. In Proceedings of the 2014 IEEE Global Humanitarian Technology Conference-South Asia Satellite (GHTC-SAS), Trivandrum, India, 26–27 September 2014.

33. Tan, Y.; Liu, W.; Qiu, Q. Adaptive power management using reinforcement learning. In Proceedings of the 2009 IEEE/ACM International Conference on Computer-Aided Design-Digest of Technical Papers, San Jose, CA, USA, 2–5 November 2009.

34. Kofinas, P.; Vouros, G.; Dounis, A.I. Energy Management in Solar Microgrid via Reinforcement Learning. In Proceedings of the 9th Hellenic Conference on Artificial Intelligence, Thessaloniki, Greece, 18–20 May 2016; pp. 1–7.

35. Kofinas, P.; Vouros, G.; Dounis, A. Energy management in solar microgrid via reinforcement learning using fuzzy reward. *Adv. Build. Energy Res.* **2017**, *12*, 1–19. [CrossRef]

36. Anvari-Moghaddam, A.; Rahimi-Kian, A.; Mirian, M.S.; Guerrero, J.M. A multi-agent based energy management solution for integrated buildings and microgrid system. *Appl. Energy* **2017**, *203*, 41–56. [CrossRef]

37. Ghorbani, S.; Rahmani, R.; Unland, R. Multi-agent Autonomous Decision Making in Smart Micro-Grids' Energy Management: A Decentralized Approach. In *Multiagent System Technologies*; Springer International Publishing: Cham, Switzerland, 2017.

38. Bogaraj, T.; Kanakaraj, J. Intelligent energy management control for independent microgrid. *Sādhanā* **2016**, *41*, 755–769.

39. Kim, H.M.; Lim, Y.; Kinoshita, T. An Intelligent Multiagent System for Autonomous Microgrid Operation. *Energies* **2012**, *5*, 3347–3362. [CrossRef]

40. Eddy, Y.S.F.; Gooi, H.B.; Chen, S.X. Multi-Agent System for Distributed Management of Microgrids. *IEEE Trans. Power Syst.* **2015**, *30*, 24–34. [CrossRef]

41. Li, Y.; Zhang, H.; Liang, X.; Huang, B. Event-Triggered-Based Distributed Cooperative Energy Management for Multienergy Systems. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2008–2022. [CrossRef]

42. Zhang, H.; Li, Y.; Gao, D.W.; Zhou, J. Distributed Optimal Energy Management for Energy Internet. *IEEE Trans. Ind. Inform.* **2017**, *13*, 3081–3097. [CrossRef]

43. Raju, L.; Sankar, S.; Milton, R.S. Distributed Optimization of Solar Micro-grid Using Multi Agent Reinforcement Learning. *Procedia Comput. Sci.* **2015**, *46*, 231–239. [CrossRef]

44. Kofinas, P.; Dounis, A.I.; Vouros, G.A. Fuzzy Q-Learning for multi-agent decentralized energy management in microgrids. *Appl. Energy* **2018**, *219*, 53–67. [CrossRef]

45. Bahramara, S.; Moghaddam, M.P.; Haghifam, M.R. Optimal planning of hybrid renewable energy systems using HOMER: A review. *Renew. Sustain. Energy Rev.* **2016**, *62*, 609–620. [CrossRef]

46. Luta, D.N.; Raji, A.K. Optimal sizing of hybrid fuel cell-supercapacitor storage system for off-grid renewable applications. *Energy* **2019**, *166*, 530–540. [CrossRef]

47. Ram, J.P.; Babu, T.S.; Rajasekar, N. A comprehensive review on solar PV maximum power point tracking techniques. *Renew. Sustain. Energy Rev.* **2017**, *67*, 826–847. [CrossRef]

48. Wei, C.; Zhang, Z.; Qiao, W.; Qu, L. Reinforcement-Learning-Based Intelligent Maximum Power Point Tracking Control for Wind Energy Conversion Systems. *IEEE Trans. Ind. Electron.* **2015**, *62*, 6360–6370. [CrossRef]