Design of Cost-Based Sizing and Energy Management Framework for Standalone Microgrid using Reinforcement Learning

Yara Khawaja[1], Issa Qiqieh[2†], Jafar Alzubi[2†], Omar Alzubi[2‡], Adib Allahham[3], and Damian Giaouris[3]

[1] *Faculty of Engineering and Technology, Applied Science Private University, Amman, Jordan*
[2†] *Faculty of Engineering, Al-Balqa Applied University, Al-Salt, Jordan,*
[2‡] *Prince Abdullah bin Ghazi Faculty of Information and Communication Technology, Al-Balqa Applied University, Al-Salt, Jordan,*
[3] *School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, UK*
Emails: [1] *y_khawaja@asu.edu.jo,* [2] {*i.qiqieh, j.zubi, o.zubi*}*@bau.edu.jo,* [3] {*adib.allahham, damian.giaouris*}*@newcastle.ac.uk*

Abstract

The standalone photovoltaic-battery energy storage (PV-BES) microgrid has gained substantial interest recently due to its ability to provide uninterrupted power to consumers in remote areas. In such microgrids, components must be precisely sized and energy must be supplied most cost-effectively at all times. This paper presents a cost-based framework for determining the optimal size and energy management of standalone microgrids using reinforcement learning. Fundamental to this framework is two essential phases; the first is finding the best size of PV-BES using an analytical and economic sizing (AES) model based on minimum levelized cost of energy (LCOE). The AES phase is then followed by optimizing the energy management strategy (EMS) of the microgrid using reinforcement learning to provide optimum cost savings. The novelty in this work can be outlined as optimizing both the size and EMS of a standalone PV-BES microgrid using the AES model and Q-learning in an integrated framework. This can lead to improved performance demonstrated in reducing the LCOE, decreasing diesel generator working hours, and enhancing PV utilization and system efficiency. The results show an advantageous reduction in total cost while meeting load requirements. Additionally, the proposed framework is evaluated using several metrics to measure the impact of employing Q-learning against the AES-finite automata model. For instance, a decrease of 22% in diesel generator working hours and an increase of 6% in PV utilization while a reduction of 11% in the LCOE is accomplished. On the other hand, the proposed framework is examined against two rule-based EMSs, load following strategy (LFS) and cycle charging strategy (CCS), and outperforms these two EMSs in terms of LCOE, PV utilization, and system efficiency.

Keywords— Sizing, Energy Management, Energy Storage, Levelized Cost of Energy, Reinforcement Learning, Q-learning.

1 Introduction

Alternative energy resources have been increasingly popular in recent decades as a means of combating the environmental damage caused by reliance on fossil fuels. In modern power systems, PV has become one of the leading renewable energy resources (RERs) since it is freely available, inexhaustible, and environmentally friendly [1, 2]. However, the PV systems are weather-dependent as any change in sunlight will have a noticeable effect on the generated energy. This is can be resolved by integrating energy storage systems (ESSs) in the microgrid to store surplus PV energy and use it as needed. Moreover, the ESS can assist in managing any challenges arising from the integration of RERs, as well as maintaining the balance between produced and consumed energy [3, 4]. A standalone microgrid consisting of PV and BES is considered one of the simplest forms of microgrids that play an important role in remote areas [5, 6]. Despite the considerable benefits of PV-BES microgrids, they face several challenges in terms of the required PV/BES capacity, and energy management of the microgrid [7]. To achieve a cost-effective system, satisfy the load, and minimize power losses, the optimal size of the microgrid components must be carefully allocated [8]. Also, constructing a robust EMS is essential to provide a reliable connection among all the components, and ensure optimal exploitation of the RERs [9, 10]. To this end, finding the optimal size-EMS combination is a persistent need for any power system to attain the most desirable benefits in terms of operation and cost. However, choosing suitable methods to accomplish this is not a simple task, especially with the presence of uncertainty in RERs and load.

Through the literature, there are many studies that have addressed the problem of finding the optimal size-EMS combination. These studies have spanned through various categories, such as analytical methods, evolutionary algorithms, and machine learning algorithms. Meta-heuristic algorithms have been utilized to determine the optimal size and placement of PVs and BESs [11]. While other works applied analytical methods to determine the penetration level of PV units in a distribution network [12].

Machine learning (ML) algorithms have recently attracted researchers in the domain of power systems, and have been widely applied to solve the complexities that appeared with

Nomenclature

α	learning rate
β	temperature coefficient of solar cell efficiency, 1/°C
η_{ch}	battery charge efficiency
η_{dch}	battery discharge efficiency
η_{inv}	inverter efficiency
η_m	PV module efficiency
η_{pv}	PV overall efficiency
η_{rt}	Round trip efficiency
η_{sys}	system overall efficiency
η_{temp}	PV temperature efficiency
γ	discount factor
A, B	diesel generator consumption curve coefficients, L/kWh
A_{pv}	PV total area, m^2
a_t	action performed at time t
BES_C	battery capacity, kWh
BHA	battery hours of autonomy, hrs
C_{ch}	cost of charging battery energy system, $
C_{dis}	cost of discharging battery energy system, $
C_{DSL}	total cost of diesel generator, $
C_{fuel}	diesel generator total fuel cost,$/L
$C_{in,BES}$	initial cost for battery, $/kWh
$C_{in,DSL}$	initial cost for diesel generator, $/kW
$C_{in,PV}$	initial cost for PV, $/kW
$C_{microgrid}$	total cost of the microgrid, $
$C_{om,BES}$	O&M cost for battery, $/kWh
$C_{om,DSL}$	O&M cost for diesel generator, $/kW
$C_{om,PV}$	O&M cost for PV, $/kW
$C_{op,BES}$	hourly operating cost of battery energy system, $
$C_{rep,BES}$	replacement cost for battery, $/kWh
$C_{rep,DSL}$	replacement cost for diesel generator, $/kW
DEG_{BES}	battery energy system degradation rate
DEG_{pv}	PV degradation rate
E_{BES}	energy produced by the BES, kWh
E_{ch}	energy used to charge the BES, kWh
E_{dis}	energy drawn from the BES, kWh
E_{DSL}	energy produced by the diesel generator, kWh
$E_{microgrid}$	total energy generated by the microgrid, kWh
E_{pv}	energy produced by PV, kWh
F_{con}	diesel generator fuel consumption, L
f_u	fuel unit cost, $/L
H	yearly module reference in-plane radiation, kW/m^2
I_{NOCT}	solar radiation at NOCT, W/m^2
I_{pv}	solar radiation, kW/m^2
j	index of year
$L_{DSL,h}$	life time of diesel generators, hrs
$L_{DSL,y}$	calculated
L_{DSL}	life time of diesel generators, years
$LCOE$	levelized cost of energy, $/kWh
M_{DSL}	diesel generator margin coefficient
N	system lifetime, years
$NOCT$	normal operating cell temperature, °C
$P_{DSL}(t)$	hourly generated power by diesel generator, kW
$P_{input}(t)$	sum of input power to battery at a specific hour, kW
$P_{L,av}$	average hourly load, kW
$P_{L,max}$	maximum load, kW
$P_L(t)$	hourly load, kW
$P_{PV,surplus}$	surplus power generated from PV
$P_{pv}(t)$	hourly power generated by PV, kW
$P_{R,DSL}$	diesel generator rated power, kW
$P_{R,pv}$	PV rated power, kW
$Q(s_{t+1}, a_t)$	Q-value of the next state s_{t+1} when performing the action a_t
$Q(s_t, a_t)$	Q-value of the current state s_t when performing the action a_t
r	discount rate
$R(s_t, a_t)$	reward given to the agent at state s_t when performing the action a_t
s_t	state of the system at time t
soc	battery state of charge
soc_{DSL}	battery soc for diesel generator operation
soc_{max}	maximum battery state of charge
soc_{min}	minimum battery state of charge
t	index of hours in a year
T_a	ambient temperature, °C
T_C	temperature of Pv cell, °C
T_{ref}	PV cell reference temperature, °C
WH_{DSL}	yearly working hours of diesel generator, hrs
Z_{DSL}	binary number for controlling diesel generator operation
AES	analytical and economic sizing
BES	battery energy system
DOD	battery depth of discharge
DSL	diesel generator
EMS	energy management strategy
ESS	energy storage system
PV	photovoltaic
RER	renewable energy resource
RL	reinforcement learning
SOH	battery state of health

the inclusion of RERs [13, 14]. Reinforcement learning (RL) is a promising scope of ML which allows an agent to learn solving problems within its environment [15, 16]. RL has shown a great ability to optimize the behavior of the components in the microgrid leading to excellent performance [17, 16]. Q-learning is a popular RL approach for solving sequential decision-making problems [18, 19]. Q-learning is an off-policy algorithm that works in real-time data management systems as it doesn't require prior knowledge of rewards or state transition probability. [20, 21].

Several researchers have utilized Q-learning to optimize microgrid EMS, for instance, Foruzan et al. [22] employed

an adaptive energy management method based on a multi-agent Q-learning approach. The microgrid reduced its reliance on the main grid and thus RERs utilization increased while the operational cost decreased. The authors in [23] incorporated a Q-learning algorithm to optimize battery scheduling in a microgrid. A number of factors were considered, including battery charge and discharge efficiency and also, the nonlinear behavior of the microgrid due to inverter efficiency. Nyong-Bassey et al. [24] presented Model-based reinforcement learning to improve the reliability of intermittent RERs by combining Power Pinch Analysis (PoPA) with several storage technologies. In [25], a real-time incentive-based demand response program was created using Q-learning. The algorithm aided the service provider in purchasing energy from its customers in order to balance load and supply while improving grid reliability. Nakabi and Toivanen [26] suggested a new grid-connected microgrid architecture including a wind generator, ESS, and a collection of thermostatically controlled and price-responsive loads. The EMS was designed to coordinate different energy sources, where several scenarios were explored using deep Q-learning methods. A multi-agent energy management approach in a grid-connected microgrid with a decentralized operation was presented in [27]. Each microgrid component was developed as an autonomous agent that uses a model-free Q-learning approach to optimize its behavior. Shang et al. [28] developed an EMS for minimizing microgrid operational costs while accounting for battery degradation. A combination of Q-learning and Monte-Carlo Tree Search was used to optimize the microgrid.

The primary idea of the preceding studies is to harness RL capabilities to optimize the EMS for different microgrid structures, which is also the main aim of our work. However, the above literature addressed enhancing the EMS for certain sizes of microgrids. This may happen because these studies dealt with existing or presumed assets of the microgrid. In this paper, we target (for the first time) optimizing both sizing and EMS for a standalone microgrid using RL. Therefore, a framework for determining the optimal size-EMS for a standalone microgrid is proposed. Firstly, an analytical and economic sizing model is employed to obtain the optimal size of the PV and BES for the standalone microgrid. Then, an RL algorithm (Q-learning) is implemented to explore the best actions leading to optimal EMS. This can be exploited to achieve additional cost reductions while ensuring load satisfaction. Inspired by our previous work [29], we present the following **new key contributions**:

- we employ and modify an analytical and economic sizing (AES) model to determine the optimal size of PV and BES in a standalone microgrid.
- we propose an integrated framework that incorporates the AES model along with the Q-learning method to optimize both size and energy management strategy (optimal size-EMS combination).
- we investigate the impact of the Q-learning method on the proposed framework for achieving substantial improvements in terms of system efficiency, PV utilization, and cost savings.
- we validate the proposed framework against Automata and two rule-based methods (load following and cycle charging strategies) and illustrate the efficiency of using AES together with Q-learning on various evaluation metrics.

The rest of the paper is structured as follows. Section 2 demonstrates the structure of the proposed microgrid with the analytical and economic descriptions for PV, BES, and DSL. Section 3 describes the proposed framework showing the optimal sizing using AES and the leveraging of the reinforcement learning algorithm to construct the optimal EMS. Then, the experimental results and discussion are presented in Section 4. Section 5 evaluates the proposed Q-learning framework against the AES-Automata approach and also against two rule-based methods. Finally, Section 6 concludes the paper.

2 Analytical and Economic Model

The Analytical and Economic Sizing (AES) model performed in this work is inspired by our previous work presented in [30] for grid-connected PV-BES system, and then modified in [29] to be applied for a standalone PV-BES-DSL-Hydrogen system. For better applicability and cost-effectiveness, we have modified the AES model to serve a new standalone structure. This section introduces the structure of the proposed standalone microgrid and illustrates the analytical models used for sizing each component in the microgrid. Also, the levelized cost of energy (LCOE) used for obtaining the optimal size is discussed in this section.

2.1 Microgrid Structure

In this paper, we have used a standalone microgrid suitable for rural areas. The microgrid depicted in Fig. 1, is equipped with the PV system, BES, DSL, DC/AC inverter, and charge controller together with the needed connections between these components.

Firstly, the allocated PV system is responsible for covering the current load, and if the load is satisfied, the excess PV energy is then stored in the BES. Such a scenario might happen during sunny periods when the PV system is capable to fulfill the required load and increase the level of stored energy in the BES. This may not be the case at other times, *i.e.*, when the generated PV energy is insufficient to cover the load, such as weather conditions or night times. On the other hand, if the stored energy is sufficient to satisfy the load, the BES is discharged until reaching the minimum specified level. The BES here is considered the first backup for the proposed standalone microgrid. It is continuously checked for ensuring the efficiency of the proposed system. The charge controller prevents the BES from being over-charged or over-discharged and thus protects it from aging. Since the load consumes AC power, the AC/DC inverter must be utilized to convert the DC power generated by the PV to AC power. Furthermore, the DSL is required in the microgrid in situations where the PV-BES system is unable to meet the load. Therefore, the DSL acts as a second backup to supply the load. In a standalone microgrid, the

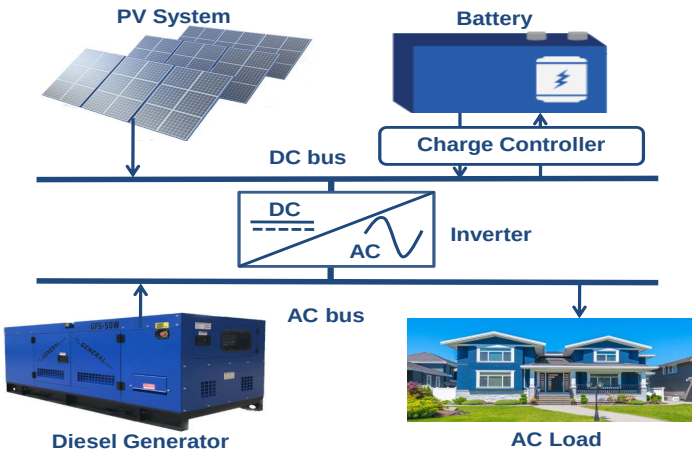

Figure 1: Schematic diagram of PV/BES/DSL microgrid

DSL is an elementary component that converts the chemical energy in the fuel into usable electrical energy to be consumed by the load. All the components in the standalone microgrid must be allocated carefully, as well as, a robust EMS has to be designed to guarantee the harmony of system operations in the microgrid. The next section presents the model used for sizing the assets in the microgrid, while the proposed EMS is explained in the next sections.

2.2 Analytical and Economic Model

This part introduces the analytical and economic (AES) model for allocating the best size of the components in the above-mentioned microgrid structure. The main aim of the proposed AES model is to specify the minimum resources required to maintain the balance of the microgrid. The AES model exercises a group of analytical equations to calculate the hourly power values produced/consumed by the microgrid components, such as the power values produced by PV/DSL, stored in the BES, and consumed by the load. These power values combined with cost specifications are then used to calculate the minimum LCOE for the microgrid. PV, DSL, and BES analytical models together with LCOE, are described below.

A. PV Model

The PV power production P_{pv} can be calculated using a set of connected equations from Eq.(1) to Eq.(4). P_{pv} can be defined as the multiplication of solar irradiance I_{pv}, PV system efficiency η_{pv}, and PV panels area A_{pv}. As in Eq.(2), η_{pv} can be affected by several parameters, such as module efficiency η_a, temperature efficiency η_{temp}, inverter efficiency η_{inv}, PV system degradation DEG_{pv}, and project lifetime N. η_{temp} is the only time-varying parameter that has an impact on η_{pv} which can be described using Eq.(3) and Eq.(4). It should be noted that the description of all required parameters used to implement the PV model is listed in Table 1.

$$P_{pv}(t) = \begin{cases} I_{pv}(t) \cdot A_{pv} \cdot \eta_{pv}(t), & I_{pv}(t) \geq 0 \\ 0, & I_{pv}(t) \leq 0 \end{cases} \quad (1)$$

$$\eta_{pv}(t) = \eta_a \cdot \eta_{temp}(t) \cdot \eta_{inv} \cdot (1 - (N-1)DEG_{pv}) \ . \quad (2)$$

$$\eta_{temp}(t) = [1 - \beta(T_C(t) - T_{ref})] \ . \quad (3)$$

$$T_C(t) = T_a(t) + [(NOCT - 20)/800] \cdot I_{pv}(t) \ . \quad (4)$$

Table 1: PV analytical model Parameters [29]

Parameter	Description	Value
η_a	PV module efficiency	14%
DEG_{pv}	PV degradation	0.5%
T_{ref}	Cell reference temperature	20°C
NOCT	Normal operating cell temperature	45°C
T_a	Ambient temperature of NOCT	20°C
I_{NOCT}	Solar irradiance at NOCT	800 W/m^2
β	Temperature coefficient of solar cell efficiency	0.005 1/°C
H	Yearly module reference in-plane radiation	$1000W/m^2$
N	PV lifetime	20 years

B. Diesel Generator Model

For standalone microgrids, DSLs are used to provide a constant supply of power to the load. If the power provided by the BES is inadequate to supply the load, the lack of power is covered by the DSL. Any excess DSL power remaining after the load has been met will be used to charge the BES. To ensure load satisfaction, the size of the DSL is determined based on the maximum load $P_{L,max}$. According to the load profiles used in this study, $P_{L,max}$ is $26.6kW$. The size of the DSL is expressed by the following equation:

$$P_{R,DSL} = M_{DSL} \cdot P_{L,max} \ . \quad (5)$$

where M_{DSL} stands for the DSL margin safety coefficient and is assumed to be 1.2. According to Eq.(5), $P_{R,DSL}$=32 kW. The hourly output of DSL is illustrated in Eq.(6) and subjected to the constraint in Eq.(7):

$$P_{DSL}(t) = \begin{cases} Z_{DSL}(t) \cdot P_{R,DSL}, & B_{DSL}(t) = 1 \\ 0, & Z_{DSL}(t) = 0 \end{cases} \quad (6)$$

$$0 \leq P_{DSL}(t) \leq P_{R,DSL} \ . \quad (7)$$

The operation of DSL is controlled by Z_{DSL} which is considered as a binary number that describes the state of the DSL at a specific hour [31]. The condition for DSL activation is linked with soc, such that if $soc(t) \leq soc_{DSL}$, the DSL operates at full capacity, where $soc_{DSL} = 25\%$.

C. Battery Energy System Model

In a standalone microgrid, the BES enables a greater amount of PV energy to be integrated into the grid, and thus increasing PV utilization. For the purpose of minimizing the operational cost of the microgrid, determining the BES capacity is essential [32]. Eq.(8) can be used to compute the BES capacity BES_C, which is expressed by BES hours of autonomy BHA and the average hourly demand $P_{L,av}$ [29]. $P_{L,av}$ is determined according to the load profiles which is used in this study and found to be 10.7 kW (for

more details regarding load profiles, see Section 4). Three values of BHA—12, 24, and 36 hours—are taken into consideration in this work to determine the best size of the BES. Selecting higher values for BHA will result in a higher capacity of the BES, which will accordingly increase the cost. On the other hand, lower values will result in a decreased capacity of the BES, which may be unable to serve the load when PV energy is unavailable.

$$BES_C = \frac{BHA \cdot P_{L,av}}{\eta_{inv} \cdot \eta_{ch} \cdot DOD} \ . \tag{8}$$

$$soc(t) = \begin{cases} soc(t-1) + \dfrac{[P_{input}(t) - P_{output}(t)] \cdot \eta_{ch}}{\eta_{inv} \cdot \ BES_C}, & P_{input}(t) > P_L(t) \\ soc(t-1) - \dfrac{P_{output}(t) - P_{input}(t)}{\eta_{inv} \cdot \eta_{dch} \cdot BES_C}, & P_{input}(t) \leq P_L(t) \end{cases} \tag{9}$$

$$P_{input}(t) = P_{pv}(t) + P_{DSL}(t) \ . \tag{10}$$

$$P_{output}(t) = P_L(t) \ . \tag{11}$$

The state of charge of BES ($soc(t)$) reflects the available capacity in the BES at a certain hour. $soc(t)$ is hourly calculated using Eq.(9). $P_{input}(t)$ is the total input power to the BES and can be obtained using Eq.(10). In this work, $P_{output}(t)$ is assumed to be as same as $P_L(t)$. The soc should be preserved between two limit levels, soc_{max} and soc_{min} as described in the following restriction:

$$soc_{min} \leq soc(t) \leq soc_{max} \ . \tag{12}$$

The description of all parameters used in equations from Eq.(8) to Eq.(12) is presented in Table 2.

D. LCOE Model

The levelized cost of energy (LCOE) is a cost-based criterion used to compare and assess various RERs. The costs are calculated over the project lifetime, providing a more accurate economic picture of the system under consideration [29]. Generally, the LCOE is computed by dividing the overall system cost during its lifetime by the energy generated from the system during the same period. Table 3 lists all the equations used for calculating the LCOE. These equations include the total cost and total energy generated cost of the PV, DSL, and BES. The general form of LCOE is presented by Eq.(13), where the total cost of the microgrid $C_{microgrid}$ (Eq.(14)) is defined as the sum of the total costs for the PV, BES, and DSL as in Eq.(17), Eq.(19), and Eq.(25), respectively. The $E_{microgrid}$ is the denominator in the LCOE equation and can be found using Eq.(15), which represents the energy generated by PV (Eq.(18)), BES (Eq.(20)), and DSL (Eq.(26)).

Table 2: All design parameters used for finding the size and soc of the BES [29]

Parameter	Description	Value
DOD	Depth of discharge	80%
η_{ch}	Charge efficiency	80%
η_{dch}	Discharge efficiency	80%
η_{inv}	Inverter efficiency	90%
soc_{min}	Minimum state of charge	20%
soc_{max}	Maximum state of charge	100%
η_{rt}	Round trip efficiency	90%
DEG_{BES}	Battery degradation rate	0.1%
BHA	Hours of autonomy	12, 24, 36 hrs
V_{BAT}	Battery voltage	48 V

The total cost for any component in the microgrid can be written as in Eq.(16), which represents the summation of the installation cost $C_{in,system}$, the operation and maintenance cost $C_{OM,system}$, and the replacement cost $C_{rep,system}$. Table 4 illustrates the cost specifications values for the PV, BES, and DSL. For this study, the lifetime of the proposed project is 20 years. Thus, the PV system has no replacement cost since its lifetime is assumed to be 20 years. The DSL is also selected to be replaced once after 10 years, while the BES is replaced once after 12 years. Equations from Eq.(22) to Eq.(24) are used to calculate the amount of consumed fuel and its cost, as well as, the lifetime of the DSL to find its replacement cost. A and B in Eq.(22) are the coefficients of the fuel consumption curve, 0.246 and 0.08145, respectively [34].

Figure 2 shows a flowchart of the process of finding LCOE for the standalone microgrid. In the first step, all required data needed to calculate the total cost and energy cost for every component in the microgrid are collected. These data consist of fuel cost information, cost specifications, and microgrid components' energy for each hour. Next, the fuel cost, total cost, and total energy cost for every component in the microgrid are computed for the first year. This follows by finding the total annual cost and total annual energy cost of the microgrid. This includes summing up the individual annual costs and annual energy costs for each component. The previous step is repeated until the end of the project lifetime (N=20). Hence, the LCOE is found by dividing the total annual cost by the total annual energy cost of all components.

For finding the minimum LCOE, the LCOE with the aforementioned steps is iterated for multiple PV and BHA values. However, the results of finding the minimum LCOE and its effectiveness on the optimal size of microgrid components are explained in detail in Section 3.1. The next section highlights how the AES model is utilized to find the optimal size of the proposed standalone microgrid. The proposed framework is described below.

3 The Proposed Framework

The proposed framework consists of two core phases; firstly, the AES model is exercised to assign the best size of the PV and BES through initial rule-based EMS ($EMS_{initial}$). Secondly, the assigned sizes, in the previous

Table 3: Equations used for calculations of the LCOE for PV, BES, and DSL in the standalone microgrid [29]

Performance measure	Definition
Levelized cost of energy	$LCOE = \frac{Total\ System\ Costs}{Total\ load}\ (\$/\text{kWh}) = \sum_{j=0}^{N} \frac{\frac{C_{microgrid}}{(1+r)^j}}{\frac{P_L}{(1+r)^j}}$. (13)
Total cost of the microgrid	$C_{microgrid} = C_{pv} + C_{BES} + C_{DSL}$. (14)
Total energy generated by the microgrid	$E_{microgrid} = E_{pv} + E_{BES} + E_{DSL}$. (15)
Total cost of a component	$C_{system} = C_{in,system} + C_{OM,system} + C_{rep,system}$. (16)
Total cost of PV	$C_{pv} = C_{in,PV} + \frac{\sum_{j=0}^{j=N} C_{om,PV}}{(1+r)^j}$. (17)
Total energy generated by PV	$E_{PV,total} = \sum_{j=0}^{j=N} \frac{\sum_{n=0}^{n=8760} E_{pv} \cdot (1 - DEG_{pv})^j}{(1+r)^j}$. (18)
Total cost of BES	$C_{BES} = C_{in,BES} + \frac{\sum_{j=0}^{j=N} C_{om,BES}}{(1+r)^j} + \frac{\sum_{j=10} C_{rep,BES}}{(1+r)^j}$, (19)
Total energy generated by BES	$E_{BES,total} = \eta_{rt} \cdot \sum_{j=0}^{j=N} \frac{\sum_{n=0}^{n=8760} E_{ch} \cdot (1 - DEG_{BES})^j}{(1+r)^j}$. (20)
Energy used to charge BES	$E_{ch}(t) = \begin{cases} P_{input}(t) - P_{output}(t), & P_{input}(t) > P_{output}(t) \\ 0, & P_{input}(t) < P_{output}(t) \end{cases}$ (21)
Fuel consumption	$F_{con}(t) = \begin{cases} A \cdot P_{R,DSL} + B \cdot P_{DSL}(t), & P_{DSL}(t) > 0 \\ 0, & P_{DSL}(t) = 0 \end{cases}$ (22)
Cost of fuel	$C_{fuel}(t) = \begin{cases} F_{con}(t) \cdot F_u, & F_{con}(t) > 0 \\ 0, & F_{con}(t) = 0 \end{cases}$ (23)
DSL lifetime	$L_{DSL,y} = \frac{L_{DSL,h}}{WH_{DSL}}$. (24)
Cost of DSL	$C_{DSL} = C_{in,DSL} + \frac{\sum_{j=0}^{j=N} C_{om,DSL}}{(1+r)^j} + \frac{\sum_{j=L_{DSL,y}} C_{rep,DSL}}{(1+r)^j} + \frac{\sum_{j=0}^{j=N} C_{fuel}}{(1+r)^j}$. (25)
Total energy generated by DSL	$E_{DSL,total} = \sum_{j=0}^{j=N} \frac{\sum_{n=0}^{n=8760} E_{DSL}}{(1+r)^j}$. (26)

phase, are examined in the microgrid environment to optimize the EMS using a reinforcement learning algorithm. The proposed framework of the standalone microgrid is illustrated in Fig. 3.

In the first phase, the framework starts with finding the best size of the PV-BES using the AES model which is explained in the next subsection in depth. The AES model requires several sets of input data, such as load and PV profiles (discussed in the results section), as well as, the costs and specifications of the components (see Section 2.2). The AES model is based on finding the minimum LCOE of multiple generated scenarios, where each scenario has different sizes of PV ranging from 10-100 kW, with an increase of 10 kW for each iteration. Three different values of BHA; 12,

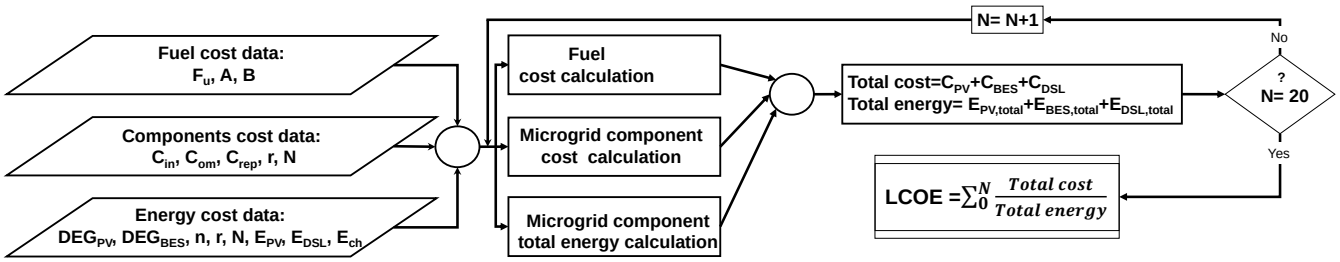

Figure 2: The process chart of calculating the levelized cost of energy (LCOE) for the standalone microgrid while lifetime N=20 years.

Table 4: Cost specifications of PV, DSL, and BES [33].

Parameter	Description	Value
$C_{in,PV}$	PV installation cost	\$1500 /kW
$C_{om,PV}$	PV O&M cost	\$30 /kW/Year
$C_{in,DSL}$	DSL installation cost	\$500 /kW
$C_{om,DSL}$	DSL O&M cost	\$0.025/hour/kW
$C_{rep,DSL}$	DSL replacement cost	$C_{in,DSL}$/10 years
$C_{in,BES}$	BES installation cost	\$213/kWh
$C_{om,BES}$	BES O&M cost	3% of $C_{in,BES}$
$C_{rep,BES}$	DSL replacement cost	$C_{in,BES}$/12 years
F_u	Fuel unit cost	1 \$/L

24, and 36 hours are also examined through the simulation. The $EMS_{initial}$ is developed from the logical and empirical rules extracted from the analytical models explained in Section 2. The LCOE is calculated for each scenario, then, the optimal size is obtained by selecting the minimum LCOE of the generated scenarios.

After determining the best size of the microgrid components, an EMS is implemented using a Q-learning algorithm which is referred as the second phase. At the beginning of this phase, the Q-table is initialized to zeros and the reward function is defined. The Q-table is updated for every episode where each episode counts as one day. The agent starts exploring the microgrid environment and based on the current state, the agent decides to move to one of the defined states by selecting a specific action. At the end of the learning process, the final Q-table is generated showing the Q-values for each state-action pair. Depending on the resultant Q-values, the selection of the best actions takes place leading to obtaining the optimal EMS. Eventually, the proposed framework can create the optimal size-EMS combination which is the main purpose of this research. Each step of this framework is fully explained in the following subsections.

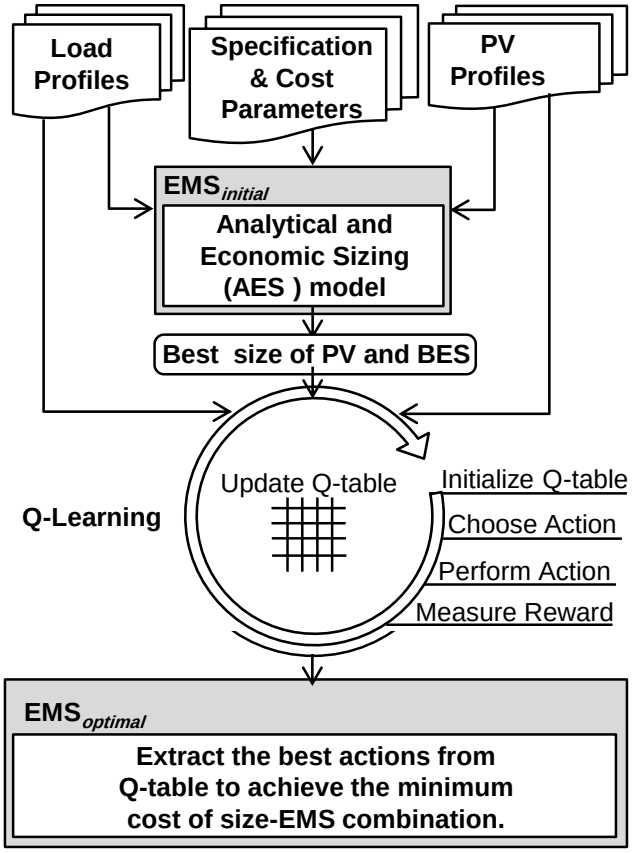

Figure 3: The proposed Framework consists of two phases:
1. applying the AES model to obtain the best size,
2. utilizing Q-learning in the EMS to extract the optimal actions with minimum cost.

3.1 Finding PV/BES Size using AES

AES is the first phase in the proposed framework illustrated in Fig. 3. The AES model was introduced in earlier work [30] for a grid-connected PV-BES microgrid which includes the operations of selling and buying energy to/from the grid. In this work, several alterations are used for adjusting the prior AES model to meet the requirements of the proposed standalone PV/BES/DSL microgrid. The alterations comprise various modifications on the P_{input}, P_{output}, and cost analysis equations. This section elaborates on these modifications, for instance, in equations Eq.(10) and Eq.(11), the P_{input} and P_{output} are modified to suit the proposed standalone microgrid. The part of the equation that is related to grid connectivity (buy and sell energy) is replaced by a new mathematical expression to involve the impact of generated power of the DSL. Therefore, the amount of energy used to charge the BES is affected (see Eq.(21)). Moreover, the total cost, as well as the total energy generated by the proposed microgrid, are also adjusted accordingly (see Eq.(14) and Eq.(15)).

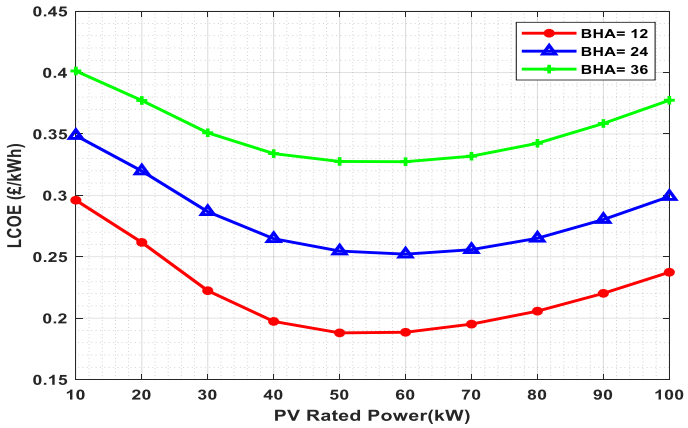

Figure 4: Levelized cost of energy for the standalone PV-BES microgrid using AES model, the PV, and BES sizes are picked at minimum LCOE.

To find the optimal size-EMS combination, the AES model should take several input data, such as load profiles, PV profiles, and cost parameters for all microgrid components. The purpose of the AES model is to produce many scenarios with various PV and BES sizes to determine the LCOE of each scenario. Fig. 4 demonstrates the calculated LCOE values for the generated scenarios. Each scenario represents one of three battery hours of autonomy (BHA) for all values of the PV rated power ranging from 10 kW to 100 kW with a step of 10 kW each time. The values of BHA used in this study are 12, 24, and 36 hours, (see Table 2). Table 5 presents the optimal size for the components of the proposed microgrid depending on the minimum LCOE of all scenarios depicted in Fig. 4. Note that, the size of the DSL is previously discussed in Section 2.2.

Allocating the sizes of the microgrid components is an essential step to obtain the maximum of these resources and avoid extra costs. However, combining the size of the microgrid with an efficient EMS provides more benefits in terms of cost and performance. The objective function and Q-learning algorithm used for optimizing the EMS are described in the following sections.

3.2 Objective Function and Constraints

In the previous subsection, we leverage the process of finding minimum LCOE to find the optimal size of the standalone microgrid. This is followed by the Q-learning phase for obtaining optimal EMS. Therefore, as the proposed framework consists of two consecutive phases, two objective functions have been put in place to control the output of each phase. The first objective function (obj_1) of microgrid sizing is to minimize the LCOE of PV/BES components while satisfying operational constraints. The second objective function (obj_2) is directed at minimizing operational costs for EMS while meeting the load at all times. Eq.(27) and Eq.(28) highlight the two objective functions of the proposed framework. The state variables of the optimization study are PV rated power, battery autonomy hours BHA, and DSL working hours WH_{DSL}. Optimizing these values should result in achieving the defined objective functions.

Table 5: The size of PV and BES in the standalone microgrid based on AES model

Subsystem	Size
PV system	50 kW
Battery energy system	218 kWh
Diesel generator	32 kW

$$obj_1 = min\ LCOE = min(\sum_{j=0}^{N} \frac{\frac{C_{microgrid}}{(1+r)^j}}{\frac{P_L}{(1+r)^j}}) \ , \tag{27}$$

$$obj_2 = min\ C_{op,ESS} = min(\frac{-DEG_{BES} * (soc(t-1) - soc(t))}{(1 - SOH_{min})}) \ , \tag{28}$$

where $C_{microgrid}$ is obtained by Eq.(14). BES state of health SOH_{min} is the ratio of remaining capacity in the BES to initial capacity (in %). SOH_{min} can be found as follows [35, 36]:

$$SOH_{min} = DEG_{BES} \cdot soc_{min} \ . \tag{29}$$

For a standalone PV/BES/DSL system, the following operational constraints should be satisfied. All these constraints have been explained in Section 2.2. Eq.(30), Eq.(31), and Eq.(32) form the applied constraints related to the sizing phase.

$$10kW \leq P_{PV,rated} \leq 100kW \tag{30}$$

$$12 \leq BHA \leq 36 \tag{31}$$

$$0 \leq P_{R,DSL} \leq M_{DSL} * P_{L,max} \tag{32}$$

Additionally, Eq.(33) and Eq.(34) represents the constraints for EMS optimizing phase using Q-learning.

$$soc_{min} \leq soc(t) \leq soc_{max} \tag{33}$$

$$P_{pv}(t) + P_{BES}(t) + P_{DSL}(t) = P_L(t) \ , \tag{34}$$

where $P_{BES}(t)$ is the hourly power drawn/stored from/in the BES.

3.3 Reinforcement Learning

RL is a branch of machine learning concerned with how an agent existing in an environment takes actions and moves to other states to gain positive rewards [15, 37]. To be more precise, a learning agent that exhibits sequential behavior over the time step t is taken into account. At each time step, the agent observes the state of the environment $s_t \in S$, and makes a decision on which action to take $a_t \in A$. Then the agent receives a reward $R(t)$ on that transition and observes the new state s_{t+1}. It is assumed that the environment is stochastic and Markovian: that the next state s_{t+1} is determined only by the current state and the current action through the state transition probability $P(s_{t+1}, s_t, a_t) = P[s_{t+1}|s_t, a_t]$. The mapping between the state and the best action is discovered through trial-and-error interaction

between the agent and its environment. This paper identifies the key features of such a learning agent and explores how it can be used to control a microgrid. The agent must therefore rely on knowledge that is generally insufficient for making a deterministic decision on the best action. When using RL to solve a real-world problem, situations like this one are common, despite knowing that the agent does not observe all available information, we proceed as though it did. To summarize, an RL agent solves a Markov Decision Problem (MDP) specified by a tuple of (S, A, P, R,) where P and R are unknown.

There are several characteristics of the microgrid including stochasticity, continuity, and partial observation. Since the set of possible actions and states is continuous, the environment is continuous. The assumption that the agent has access to the environment's actual state cannot be satisfied in this situation due to partial observability [38].

One of the most well-known algorithms in RL is Q-learning which evaluates the mapping process between states and actions using Q-function. The following section demonstrates the Q-learning and how the EMS in this proposed work is modeled using this algorithm.

A. Using Q-learning to implement EMS

Q-learning is a reinforcement learning algorithm that identifies the best action to take based on the current state of the problem and is used for sequential optimization and control in uncertain environments [39]. Q-learning is a model-free algorithm where the agent in the environment is trained to evaluate the quality of accomplished actions telling the agent which actions to perform to learn the optimal control policy. This paper considers an MDP with an optimal policy $\pi(s)$ defined as a deterministic mapping between a set of states and a set of actions. Such that there is one best action for each state or possibly several optimal actions. It is important to recognize the "value" of a state-action pair, which the agent takes the advantage of in selecting a particular action a in a particular state s for the purpose of optimizing the objective function and this value is called Q-value $Q(s_t, a_t)$ [37]. The optimal policy is denoted by π^* and its Q-value is Q^*, where $\pi^*(s) = \arg max\ Q^*(s, a)$. Q^* can be learned recursively by random approximation whose main iteration is defined by Bellman equation:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha[R(s_t, a_t) + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \tag{35}$$

Under appropriate assumptions, the series Q converges to Q^*. The agent will eventually learn and converge to the Q value of the optimal policy. Several such episodes will be repeated from the initial state to the final state for this task. The term max $Q(s_{t+1}, a)$ is the maximum of the Q-value of the next state calculated with all possible actions. In this paper, the Q-values are stored in Q-table. The initialization of this Q-table is random, then, the agent changes the Q values using Eq.(35) as it engages with its environment. Choosing the best action is very crucial, the agent must initially try several actions (explore the set of actions), and as it learns, it will increasingly concentrate on the actions that seem to be the best. There are various ways to handle action selection, the selected approach in this paper is known as the ϵ-greedy approach.

In Eq.(35), α is the learning rate that controls the rate or speed at which the agent learns [15]. The value of α is usually between 0 and 1, values closer to zero mean the agent's ability to learn is minimal and the Q-values are never updated. On the other hand, setting α to values near 1 mean the agent learns fast and the learning process occurs quickly, accordingly, in this research α is set to 0.5. $R(s_t, a_t)$ is the reward given to the agent based on the taken action. γ is the discount factor that indicates how future rewards are important to the current state [15]. The value of γ is between 0 and 1, lower values state the agent only cares about immediate rewards, while higher values express how the agent accounts for future rewards. The value of γ used in the simulations is 0.9. In this paper, the values for α and γ chosen based on multiple experiments, where Q-learning algorithm was tested over various possible α and γ values. It is found that when α is between 0.5 and 0.6 and γ between 0.7 and 0.9 the best performance of the Q-learning algorithm is obtained. Accordingly, α=0.5 and γ=0.9 are selected for the experiments performed in this paper.

The definitions of the Q-learning element are explained as follows:

- Environment: the environment in this study is the standalone microgrid which has been described in Section 2.1. The environment has several states and the agent needs to interact with the environment to collect data regarding its current state.
- Agent: the agent learns how to operate best in an environment by experience. The agent is the learner that is responsible for communicating with the environment, gathering information regarding the current state, and choosing an action to move the environment to the next state. The more the agent is involved in the learning process, the more the best decisions it can perform. In this study, the agent is the EMS that is accountable for managing the energy production by the components in the microgrid, as well as energy consumption by the load while ensuring minimum cost.
- State Space: in this research, the states are divided into 24 states, where each state represents one hour in the day. The agent needs to check the power generated by the PV and DSL, power stored or drawn by the BES, and the load power at each hour. Based on this information the agent will make a decision on which action to perform. Accordingly, the state space for this problem is S=$\{s_1, s_2, ..., s_{24}\}$.
- Action Space: the actions selected in this work are related to BES operation. Charging/DSL OFF denotes to charging the BES using surplus PV power after the load is met. While charging/DSL ON relates to charging BES while DSL is ON using the extra power after the load is satisfied. Here the DSL operates only for one hour, and consequently the next hour the agent

must decide whether to turn it on again or not depending on the BES *soc* ($soc \leq 25\%$). Finally, discharging action refers to periods when there is no PV power, nor DSL power, and the BES $soc_{min} < soc < soc_{max}$, then this action must be selected to ensure covering the load continuously. The action space is defined by A= {charging/DSL OFF, charging/DSL ON, discharging}.

- Reward: provides feedback for a policy so that it can learn desirable behavior. Positive rewards are given for desirable behaviors and negative rewards for undesirable behavior. Reward formulation is discussed later in this section.

As introduced in this section, the EMS is modeled as an individual agent which makes decisions, responds, and adapts to changes in the environment based on specific rules. The environment is constantly changing and distinguished by hourly available PV power output, load, and BES *soc*. As the agent learns from the environment, it accumulates rewards until it maximizes the total rewards received from it. Algorithm 1 demonstrates the Q-learning algorithm used to obtain the best actions leading to optimal EMS. The algorithm starts with determining the values of α, γ, and ϵ. Then an empty Q-table is initialized with 24 rows (states) and three columns (actions). For the first hour in the day when $t = 0$, the values of P_{input} and P_{output} are specified (see Eq.(10) and Eq.(11)). Based on these values, all the possible actions are identified. Then, the agent chooses an action from the action set based on the ϵ_greedy algorithm by calling the function ϵ_greedy_Action_Selection. This function returns an action a_t to be performed by the agent. The algorithm iterates for 24 hours time steps until $episode_{max}$ is reached.

B. Reward Function

The reward function is designed to direct the EMS agent to follow the optimal actions to achieve minimum operation cost for the standalone microgrid while meeting the load requirements. The reward is formulated based on maximizing the cost of charging and discharging the BES and minimizing the working hours of DSL and, as a result, fuel consumption. The operational cost is calculated hourly for every operating component. For the action a_1=charging/DSL OFF, the reward equals the BES hourly operation cost (C_{op}(t)) multiplied by the charging energy E_{ch} at that hour. Alternatively, when a_t= charging/DSL ON, the BES is charged from the remaining DSL power after the load is completely covered. It is important to note that DSL operates at 100% of its capacity, and the amount of power generated from the DSL always covers the load plus a safety margin of 10-20%. The reward function for this action is calculated by taking the negative of the operation cost of the DSL $C_{DSL,OM}$ multiplied by the hourly power generated by the DSL $P_{DSL}(t)$, and then added to the hourly operation cost of charging the BES. Finally, the reward for the discharging action a_3 is the C_{op}(t) multiplied by hourly discharging energy (E_{dis}). The value of $C_{op}(t)$ is found using Equations (36) and (30), which mainly depends on the BES state of health SOH, an important indicator of battery life and reflects the ability of

Algorithm 1 Q-Learning for EMS Optimization

Input: α: learning rate, γ: discount factor, $\epsilon \in$ [0,1]
Output: A Q-table containing $Q(S, A)$ pairs defining the estimated output policy π^*

```
Initialize Q(s_t, a_t) ∀ s_t ∈ S, ∀ a_t ∈ A(s_t) arbitrarily
for episode ≤ episode_max do
    Initialize the agent s_0, t ← 0.
    for t=0 to 23 do
        a_t ← CALL ε_greedy_Action_Selection(Q, s, ε)
        Take action a_t, observe R and s_{t+1}
        Update Q(s_t, a_t) using Eq.(35)
        s_t ← s_{t+1}
        if t=23 then
            s_t is terminal
        end if
    end for
end for
Function ε_greedy_Action_Selection(Q, s_t, ε)
k ← random number ∈ (0,1)
if k < ε then random action from A(s_t)
else
    max Q(s_t, .)
end if
return selected action a_t
End Function
```

the BES to store and deliver energy [35, 36].

$$C_{op,ESS}(t) = C_{in,BES}\frac{(-DEG_{BES} * (soc(t-1) - soc(t)))}{(1 - SOH_{min})} \quad , \tag{36}$$

The idea behind the reward function is to increase utilization of the energy stored in the BES (energy from the PV system) while lowering the number of DSL working hours (WH_{DSL}) and as a consequence, minimizing the energy drawn from the DSL. The reward function designed in this work is illustrated in Eq.(37):

$$R(s_t, a_t) = \begin{cases} C_{op,ESS}(t) \cdot E_{ch}(t), & a_t = charging/DSLOFF \\ -(C_{DSL,OM} \cdot P_{DSL}(t) - C_{fuel}(t)) & \\ +C_{op,ESS}(t) \cdot E_{ch}(t), & a_t = charging/DSL\ ON \\ C_{op,ESS}(t) \cdot E_{dis}(t), & a_t = discharging \end{cases} \tag{37}$$

After the proposed framework has been comprehensively described, the next section addresses the results obtained from this framework.

4 Experimental Results

The simulations were done on Matlab R2021a using real data profiles for both PV and load. As a case study, the hourly PV profiles are generated for Amman city, Jordan based upon National Renewable Energy Laboratory (NREL) [40]. Also, the load profiles are computed using a short-term forecasting algorithm for a typical load profile for a rural area in Amman. The algorithm is written

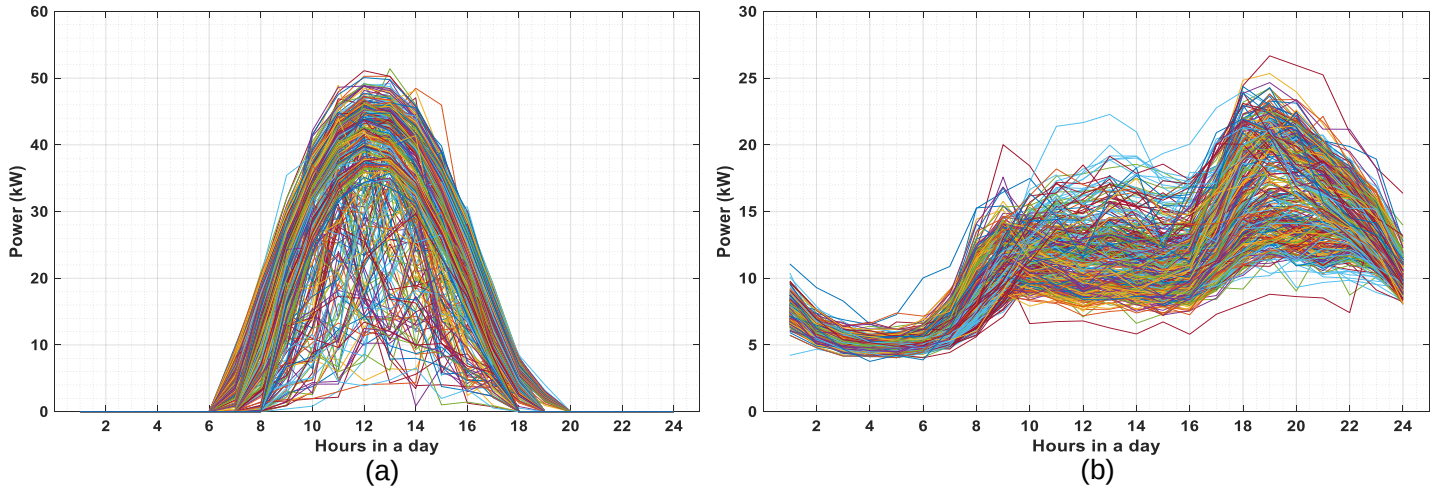

Figure 5: Data profiles used by the proposed framework divided into 24 hours for 365 days: (a) PV profiles, and (b) load profiles

by Matlab and performs an accurate computation of a year ahead for hour-by-hour electrical load based on Ammanś current load measurements. Note that the proposed framework can be applied to any area with known PV and load profiles. Fig. 5 illustrates the PV and load profiles described here and plotted for 24 hours 365 days.

Firstly, the PV and BES having an initial EMS are examined using the AES model, and the optimal size is obtained. Then, the EMS with the optimal size is implemented in the Q-learning algorithm to find the best actions that lead to optimal EMS. Finally, a validation process takes place to compare the results obtained from the AES model and Q-learning algorithm using several evaluation metrics. The dataset used for the learning process is divided into 365 days, and 24 hours each day. The length of each episode is the hours of one day, so the agent has been trained for 365 episodes.

Figure 6 demonstrates the accumulated Q-values returned by the Q-function after the learning period has been completed. The Q-function returns the Q-values calculated for every action (charging/DSL OFF, charging/DSL ON, and discharging) over 24 hours and for 365 days (number of episodes), see section 3.3). According to the figure, the time scale can be divided into five-time intervals. From midnight to 4 AM, greater Q-values are obtained for the discharging action, and the Q-values line for the charging/DSL ON are rising from zero at midnight until it reaches its maximum at 4 AM. This means that the agent learned that discharging is the best action to perform during this interval due to the availability of power stored in the BES. From 4 AM to 8 AM; the Q-values for the two actions, discharging and charging/DSL ON are very close to each other, with the latter higher. While the Q-values line for the charging/DSL OFF action is having a steep rise from zero but it did not exceed the charging/DSL ON and discharge lines. The dominant action in this interval is charging/DSL ON. After 8 AM and till 4 PM; there is a growth of the Q-values of the charging/DSL OFF action line, exceeding the other two lines; discharging and charging/DSL ON actions. At this interval, the power generated by the PV is used to cover the load and charge the BES as well. From 4 PM to 8 PM: during this interval, it can be observed from the figure the Q-values of the discharging action surpass the charging/DSL OFF action (which has a steep decline) and charging/DSL ON. This is because the PV power is decaying, while the BES is completely charged by the surplus PV power. After 8 PM and until midnight; the Q-values lines of charging/DSL ON and discharging actions are adjacent with the charging/DSL ON line exceeding the discharging line. From this figure, it can be concluded the general structure for the EMS on a typical day if is followed, an optimal operation of the microgrid is achieved with minimum LCOE.

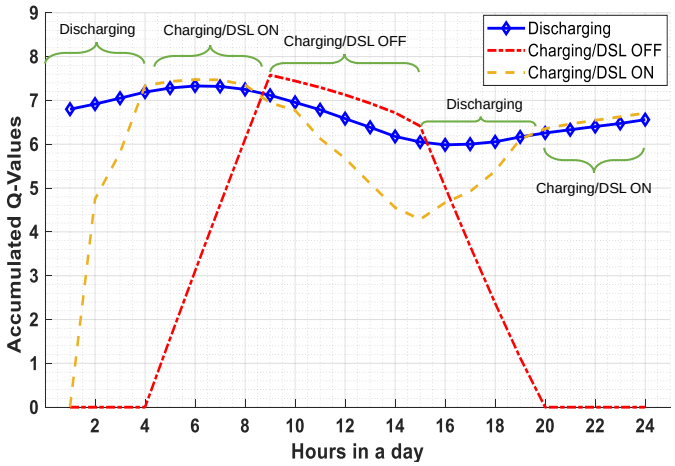

Figure 6: Accumulated Q-values for 365 days of training data showing the best actions to perform during a day to reach the optimal EMS.

Figures 7 and 8 show PV, DSL, and load power over two consecutive days at the beginning of January and June. PV produces significant amounts of energy throughout the summer and is able to cover the load entirely, with any surplus power being used to charge the BES. Accordingly, *soc* of the BES reaches a maximum value during the PV peak period as shown in Fig. 7. When the PV power drops during night hours, the BES begins discharging to cover the load instead of operating the DSL. Therefore, DSL power does not appear in Fig. 7. However, in winter, the PV power is insufficient to meet the load and charge the BES. As a

result, the DSL must operate to cover the load and the extra power is used to charge the BES. As can be seen from Fig. 8, the trend of the *soc* line is consistent with the power generated from PV and DSL and consumed by the load. Additionally, the action performed during the day from the charging, discharging, and turning on DSL follows the same rules as have been obtained by Q-learning Fig. 6.

5 Evaluation and Discussions

To illustrate the influence of harnessing reinforcement learning on the proposed framework, we evaluate the efficiency of the proposed framework against two different methods. The first compares the proposed framework against the AES-finite automata model, and then an evaluation of the proposed framework against two rule-based EMSs is demonstrated.

5.1 Evaluation Against Finite Automata Method

Modeling the EMS using finite automata has many advantages in terms of reducing the complexity of the system, simplifying the process of adding or changing the operating conditions as well as, and increasing the ability to add or remove components to the microgrid. To evaluate the proposed framework, a comparison between the proposed framework and a previous work reported in [29] is conducted. The previous work utilized finite automata together with AES to find the optimal size-EMS combination of a hybrid standalone PV/BES/DSL/hydrogen system. In this evaluation, the finite automata is employed with the AES model on the standalone PV/BES/DSL microgrid. Then, several evaluation metrics are introduced to highlight the differences between the proposed framework and the AES-finite automata model. The evaluation metrics of interest are DSL working hours WH_{DSL}, LCOE, system efficiency (η_{sys}), and PV utilization ($PV_{utilization}$). The reason for choosing these metrics is their ability to emphasize the changes that occurred after applying the proposed framework. For example, the LCOE measures the profitability of the microgrid during its lifetime, so minimizing the LCOE is essential. Fig. 9 represents a bar chart comparing the proposed framework and the AES-finite automata model. The Figure shows the reduction in WH_{DSL} from 693 hours to 537 hours when employing the proposed framework. Additionally, there are improvements in system efficiency η_{sys} and PV utilization $PV_{utilization}$. The proposed framework enhanced the $PV_{utilization}$ and η_{sys} by 6% and 3%, respectively. All the improvements whether they are gains or drops are listed in Table 6. The reason for the superiority of the proposed framework over the AES-finite automata model is that the latter explores all the possible paths and selects the best path according to the given rewards that leads to the optimal EMS with minimum operating costs. As the optimal size-EMS combination with the proposed framework has been reached, improvements in the evaluation metrics are notable.

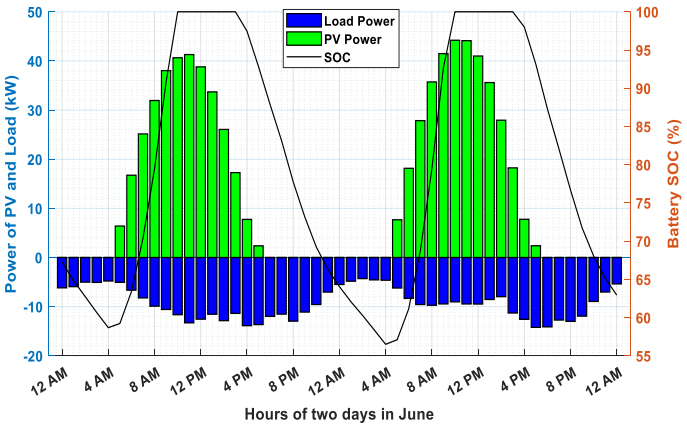

Figure 7: Power production from PV and DSL and load power consumption for random two consecutive days during June

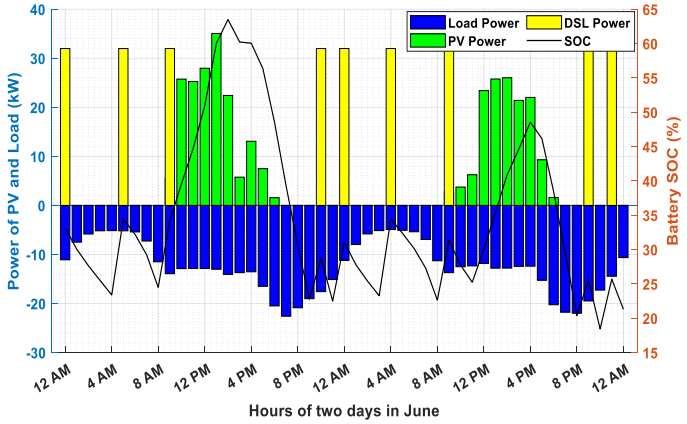

Figure 8: Power production from PV and DSL and load power consumption for random two consecutive days during January

Table 6: A comparison between the AES and Q-learning algorithm using several evaluation metrics

Evaluation Metrics	AES-FA	Framework	Improvements
$WH_{DSL}(hrs)$	693	537	22.5% reduction
$LCOE$(\$/kWh)	0.188	0.1673	11% reduction
$PV_{utilization}$	92%	98%	6% increase
η_{sys}	93%	96%	3% increase

5.2 Evaluation Against Rule-based Methods

Two rule-based methods are implemented to be evaluated

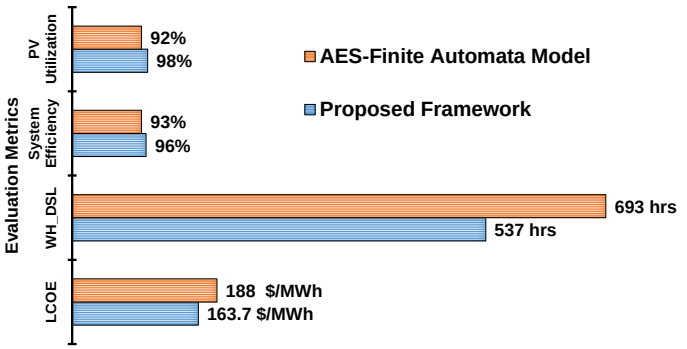

Figure 9: A bar chart shows the improvements in the evaluation metrics of the proposed framework compared to the AES-finite automata model.

against the proposed framework. These rule-based methods are load following strategy (LFS) and cycle charging strategy (CCS).

A. Comparison with Load Following Strategy (LFS)

The first rule-based EMS uses the principle of LFS, that is the DSL is only capable of providing power to cover the load. For every hour, the P_{input} and P_{output} (Equations (10) and (11)) are specified. If the $P_{input} < P_{output}$, then the load is supplied by the available energy and the BES *soc* will be checked, if *soc* is greater than 25%, the load will be supplied immediately by the BES. However, if the BES is unable to fulfill the load, the DSL produces only enough power to satisfy the load without the ability to charge the BES [41]. On the other hand, if $P_{input} > P_{output}$, then the load will be supplied by the available energy from the PV, and the surplus energy will be directed to charge the BES.

B. Comparison with Cycle Charging Strategy (CCS)

The rules used in this strategy follow the same rules as the LFS, except that if DSL is in operation, its output will be equal to the rated power. This means that the DSL will supply the load at times when PV and BES are unavailable. Any surplus energy generated from the DSL will be directed to charge the BES until it reaches soc_{max} [41].

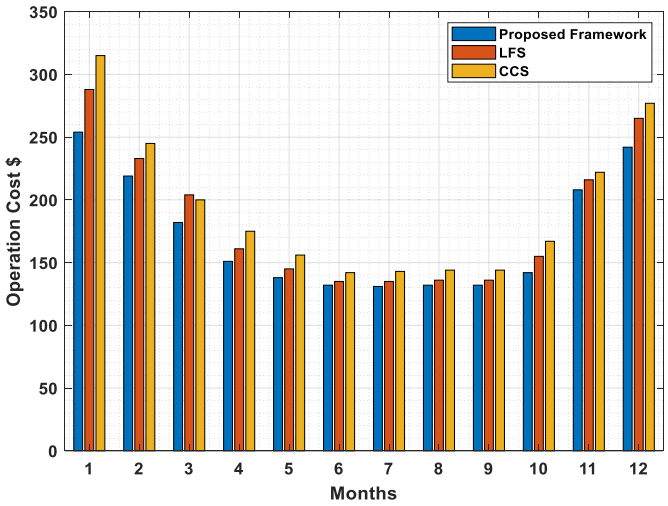

Figure 10: A bar chart illustrates the evaluation of the proposed framework against LFS and CCS

Table 7: A comparison between the AES and Q-learning algorithm using several evaluation metrics

Evaluation Metrics	Framework	LFS	CCS
$WH_{DSL}(hrs)$	537	623	**522**
$LCOE$(\$/kWh)	**0.1673**	0.1922	0.2136
$PV_{utilization}$	**98%**	96%	95%
η_{sys}	**96%**	94%	93%

Fig. 10 represents a bar chart demonstrating the operation cost for the standalone microgrid components. The chart illustrates the amount of money in $ spent on the operation cost of the microgrid per month during one year when employing the proposed framework, LFS, and CCS. It is clear that the proposed framework provides the minimum values of operation cost over all months. The CCS has the greatest operation cost values, whereas LFS has lower operation cost values than the CCS but higher than the proposed framework. It can be concluded that the proposed framework surpasses the LFS and CCS in terms of operation costs, this is due to the feature that Q-learning adds to the framework. Additionally, combining the best size for the microgrid components (using the AES model) together with the Q-leaning algorithm gives the proposed framework its capability to obtain the optimal size-EMS combination.

Table 7 lists the WH_{DSL}, $PV_{utilization}$, LCOE, and η_{sys} values for the proposed framework, LFS, and CCS. It can be inferred that the proposed framework outperforms the LFS by having smaller values in WH_{DSL}, LCOE, $PV_{utilization}$, and η_{sys} (values in bold). When comparing the proposed framework with the CCS, it can be noted that the framework exceeds the CCS by having lower values in LCOE, $PV_{utilization}$, and η_{sys}. Due to the fact that DSL uses its full power when PV or BES energy is not available, CCS has fewer values of WH_{DSL}. As a result, there will be more energy than the load needs and the surplus energy will be used to charge the BES. This rule for DSL operation has affected the WH_{DSL} values making them lower for CCS. The result in the table justifies the importance of considering the AES with Q-learning over two rule-based methods and the effectiveness of the proposed microgrid in improving the LCOE, η_{sys}, and $PV_{utilization}$.

6 Conclusion

In this paper, an optimization framework for both size and EMS of a PV/BES/DSL standalone microgrid is presented based on reinforcement learning. The proposed framework consists of two relevant phases; firstly, finding the optimal size of a PV and BES using AES. The second phase involves developing an EMS by leveraging the Q-learning algorithm to find the most promising set of actions for the given environment. Using the proposed framework a 50 kW size of the PV and 218 kWh capacity for the BES are obtained after performing the first phase. Following that, the EMS agent exploits these sizes and performs actions based on the defined states and rewards to guide the agent in selecting the actions with minimum operational costs. After the second phase, the highest Q-values of the proposed actions are aggregated for one day, such that the obtained actions represent the most efficient EMS.

To highlight the advantages of incorporating Q-learning with the AES model, an evaluation study between the proposed framework and the AES-finite automata framework is carried out. Due to Q-learning, the diesel generator working hours have been reduced by 22% compared to using AES-finite automata, which reduced the LCOE by 11%. Moreover, there is an increase in PV utilization by 6% resulting in system efficiency improvements of 3%. Additionally, two rule-based EMSs are implemented for the standalone microgrid, their evaluation metrics are compared to those for the proposed framework. The results obtained have shown an advantageous reduction in LCOE while increasing PV uti-

lization and system efficiency. We believe that reinforcement learning can be used with already existing sizing methods to extract manifold benefits in microgrids in terms of cost reduction, increased PV utilization, and reduced diesel generator working hours.

Acknowledgement

The authors would like to thank Applied Science Private University (Jordan), Al-Balqaa Applied University (Jordan), and Newcastle University (UK) for their funding and support.

References

[1] S. Yin, J. Wang, Z. Li, and X. Fang, "State-of-the-art short-term electricity market operation with solar generation: A review," *Renewable and Sustainable Energy Reviews*, vol. 138, p. 110647, 2021.

[2] M. Gul, Y. Kotak, and T. Muneer, "Review on recent trend of solar photovoltaic technology," *Energy Exploration & Exploitation*, vol. 34, no. 4, pp. 485–526, 2016.

[3] D. Giaouris, A. I. Papadopoulos, C. Ziogou, D. Ipsakis, S. Voutetakis, S. Papadopoulou, P. Seferlis, F. Stergiopoulos, and C. Elmasides, "Performance investigation of a hybrid renewable power generation and storage system using systemic power management models," *Energy*, vol. 61, pp. 621 – 635, 2013.

[4] A. Z. AL Shaqsi, K. Sopian, and A. Al-Hinai, "Review of energy storage services, applications, limitations, and benefits," *Energy Reports*, vol. 6, pp. 288–306, 2020. SI:Energy Storage - driving towards a clean energy future.

[5] B. Zhao, X. Zhang, J. Chen, C. Wang, and L. Guo, "Operation optimization of standalone microgrids considering lifetime characteristics of battery energy storage system," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 4, pp. 934–943, 2013.

[6] R. Rahmani, I. Moser, and M. Seyedmahmoudian, "Multi-agent based operational cost and inconvenience optimization of pv-based microgrid," *Solar Energy*, vol. 150, pp. 177–191, 2017.

[7] V. Murty and A. Kumar, "Multi-objective energy management in microgrids with hybrid energy sources and battery energy storage systems," *Protection and Control of Modern Power Systems*, vol. 5, no. 1, pp. 1–20, 2020.

[8] Q. Ma, X. Huang, F. Wang, C. Xu, R. Babaei, and H. Ahmadian, "Optimal sizing and feasibility analysis of grid-isolated renewable hybrid microgrids: Effects of energy management controllers," *Energy*, vol. 240, p. 122503, 2022.

[9] M. A. Ramli, H. Bouchekara, and A. S. Alghamdi, "Efficient Energy Management in a Microgrid with Intermittent Renewable Energy and Storage Sources," *Sustainability*, vol. 11, pp. 1–28, July 2019.

[10] S. Ouédraogo, G. A. Faggianelli, G. Pigelet, G. Notton, and J. L. Duchaud, "Performances of energy management strategies for a photovoltaic/battery microgrid considering battery degradation," *Solar Energy*, vol. 230, pp. 654–665, 2021.

[11] C. D. Rodríguez-Gallegos, D. Yang, O. Gandhi, M. Bieri, T. Reindl, and S. Panda, "A multi-objective and robust optimization approach for sizing and placement of pv and batteries in off-grid systems fully operated by diesel generators: An indonesian case study," *Energy*, vol. 160, pp. 410–429, 2018.

[12] D. Q. Hung, N. Mithulananthan, and K. Y. Lee, "Determining pv penetration for distribution systems with time-varying load models," *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 3048–3057, 2014.

[13] A. Perera and P. Kamalaruban, "Applications of reinforcement learning in energy systems," *Renewable and Sustainable Energy Reviews*, vol. 137, p. 110618, 2021.

[14] S. M. Miraftabzadeh, F. Foiadelli, M. Longo, and M. Pasetti, "A survey of machine learning applications for power system analytics," in *2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I CPS Europe)*, pp. 1–5, 2019.

[15] R. S. Sutton and A. G. Barto, "Reinforcement learning: An introduction second edition," 2018.

[16] A. Perera and P. Kamalaruban, "Applications of reinforcement learning in energy systems," *Renewable and Sustainable Energy Reviews*, vol. 137, p. 110618, 2021.

[17] M. Glavic, R. Fonteneau, and D. Ernst, "Reinforcement learning for electric power system decision and control: Past considerations and perspectives," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 6918–6927, 2017. 20th IFAC World Congress.

[18] C. J. C. H. Watkins and P. Dayan, "Technical note q-learning.," *Mach. Learn.*, vol. 8, pp. 279–292, 1992.

[19] C. Liu and Y. L. Murphey, "Optimal power management based on q-learning and neuro-dynamic programming for plug-in hybrid electric vehicles," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 6, pp. 1942–1954, 2019.

[20] B. Jang, M. Kim, G. Harerimana, and J. W. Kim, "Q-learning algorithms: A comprehensive classification and applications," *IEEE Access*, vol. 7, pp. 133653–133667, 2019.

[21] X. Lin, S. Zeng, and X. Li, "Online correction predictive energy management strategy using the q-learning based swarm optimization with fuzzy neural network," *Energy*, vol. 223, p. 120071, 2021.

[22] E. Foruzan, L.-K. Soh, and S. Asgarpoor, "Reinforcement learning approach for optimal distributed energy management in a microgrid," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5749–5758, 2018.

[23] B. V. Mbuwir, F. Ruelens, F. Spiessens, and G. Deconinck, "Battery energy management in a microgrid using batch reinforcement learning," *Energies*, vol. 10, no. 11, p. 1846, 2017.

[24] B. E. Nyong-Bassey, D. Giaouris, C. Patsios, S. Papadopoulou, A. I. Papadopoulos, S. Walker, S. Voutetakis, P. Seferlis, and S. Gadoue, "Reinforcement learning based adaptive power pinch analysis for energy management of stand-alone hybrid energy storage systems considering uncertainty," *Energy*, vol. 193, p. 116622, 2020.

[25] R. Lu and S. H. Hong, "Incentive-based demand response for smart grid with reinforcement learning and deep neural network," *Applied Energy*, vol. 236, pp. 937–949, 2019.

[26] T. A. Nakabi and P. Toivanen, "Deep reinforcement learning for energy management in a microgrid with flexible demand," *Sustainable Energy, Grids and Networks*, vol. 25, p. 100413, 2021.

[27] E. Samadi, A. Badri, and R. Ebrahimpour, "Decentralized multi-agent based energy management of microgrid using reinforcement learning," *International Journal of Electrical Power & Energy Systems*, vol. 122, p. 106211, 2020.

[28] Y. Shang, W. Wu, J. Guo, Z. Ma, W. Sheng, Z. Lv, and C. Fu, "Stochastic dispatch of energy storage in microgrids: An augmented reinforcement learning approach," *Applied Energy*, vol. 261, p. 114423, 2020.

[29] Y. Khawaja, A. Allahham, D. Giaouris, C. Patsios, S. Walker, and I. Qiqieh, "An integrated framework for sizing and energy management of hybrid energy systems using finite automata," *Applied Energy*, vol. 250, pp. 257–272, 2019.

[30] Y. Khawaja, D. Giaouris, H. Patsios, and M. Dahidah, "Optimal cost-based model for sizing grid-connected pv and battery energy system," in *2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, pp. 1–6, 2017.

[31] D. Giaouris, A. I. Papadopoulos, P. Seferlis, S. Voutetakis, and S. Papadopoulou, "Power grand composite curves shaping for adaptive energy management of hybrid microgrids," *Renewable Energy*, vol. 95, pp. 433 – 448, 2016.

[32] F. Al-Turjman and N. Gowthaman, *Advanced controllers for smart cities : an industry 4.0 perspective*. Springer, 2020.

[33] A. F. Altun and M. Kilic, "Design and performance evaluation based on economics and environmental impact of a pv/wind/diesel and battery standalone power system for various climates in turkey," *Renewable Energy*, vol. 157, pp. 424–443, 2020.

[34] P. E. Campana, L. Wästhage, W. Nookuea, Y. Tan, and J. Yan, "Optimization and assessment of floating and floating-tracking pv systems integrated in on- and off-grid hybrid energy systems," *Solar Energy*, vol. 177, pp. 782 – 795, 2019.

[35] E. Vanem, C. B. Salucci, A. Bakdi, and Øystein Å sheim Alnes, "Data-driven state of health modelling—a review of state of the art and reflections on applications for maritime battery systems," *Journal of Energy Storage*, vol. 43, p. 103158, 2021.

[36] F. Ramahatana and M. David, "Economic optimization of micro-grid operations by dynamic programming with real energy forecast," *Journal of Physics: Conference Series*, vol. 1343, p. 012067, nov 2019.

[37] P. Kofinas, A. Dounis, and G. Vouros, "Fuzzy q-learning for multi-agent decentralized energy management in microgrids," *Applied Energy*, vol. 219, pp. 53–67, 2018.

[38] T. Levent, P. Preux, E. le Pennec, J. Badosa, G. Henri, and Y. Bonnassieux, "Energy management for microgrids: a reinforcement learning approach," in *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pp. 1–5, 2019.

[39] B. Jang, M. Kim, G. Harerimana, and J. W. Kim, "Q-learning algorithms: A comprehensive classification and applications," *IEEE Access*, vol. 7, pp. 133653–133667, 2019.

[40] NREL, "Pv watts calculator." http://pvwatts.nrel.gov/pvwatts.php. Accessed 10.01.2021.

[41] A. Chauhan, S. Upadhyay, M. Khan, S. Hussain, T. S. Ustun, *et al.*, "Performance investigation of a solar photovoltaic/diesel generator based hybrid system with cycle charging strategy using bbo algorithm," *Sustainability*, vol. 13, no. 14, p. 8048, 2021.