International Journal of Powertrains

ISSN online: 1742-4275 - ISSN print: 1742-4267
https://www.inderscience.com/ijpt

Deep-Q-network-based energy management of multi-resources in limited power micro-grid

Nabil Jalil Aklo, Mofeed Turky Rashid

DOI: 10.1504/IJPT.2023.10049505

Article History:

Received:	17 December 2021
Last revised:	09 March 2022
Accepted:	26 March 2022
Published online:	20 March 2023

Deep-Q-network-based energy management of multi-resources in limited power micro-grid

Nabil Jalil Aklo*

Electrical Engineering Department,
University of Basrah,
Basrah, Iraq
and
Electrical and Electronic Engineering Department,
University of Thi-Qar,
Thi Qar, Iraq
Email: nabilj.aklo@utq.edu.iq
*Corresponding author

Mofeed Turky Rashid

Electrical Engineering Department,
University of Basrah,
Basrah, Iraq
Email: mofeed.rashid@uobasrah.edu.iq

Abstract: To overcome the shortage of power supply to the rural area, a hybrid connected mode micro-grid (MG) is proposed. It is suggested to include a diesel generator (DG) and renewable energy resources (RER) with a limited power of utility grid. To ensure the availability of fuel supply, the take-or-pay method is employed. In this paper, a smart energy management system (EMS) has been proposed to control the operation of hybrid MG, in addition to ensuring complete fuel disbursement under the scheduling of fuel supply. To facilitate the construction of EMS, a free model-based reinforcement learning (RL) algorithm has been employed for this purpose, in which the design of this algorithm depends on deep Q-network (DQN). The simulation of the algorithm has been achieved by MATLAB to validate the proposed system; the results showed a good performance of the technique compared with the performance achieved by improved particle swarm optimisation (IPSO) algorithm.

Keywords: energy management system; renewable energy resources; reinforcement learning; deep Q-learning; take-or-pay.

Reference to this paper should be made as follows: Aklo, N.J. and Rashid, M.T. (2023) 'Deep-Q-network-based energy management of multi-resources in limited power micro-grid', *Int. J. Powertrains*, Vol. 12, No. 1, pp.25–53.

Biographical notes: Nabil Jalil Aklo graduated with first-class honours with an Engineer degree in Control and Systems Engineering from the University of Technology, Baghdad, in 2006. He received his Master's in Electrical Engineering (Control) from the Faculty of Electrical Engineering, University Basrah, in 2012. Currently, he is PhD student in the Faculty of Electrical

Engineering, University Basrah. He has been a faculty member since 2012 at the University of Thi-Qar in Electrical Engineering Factuality. His research interests include control strategies for automation, renewable energy management and artificial intelligent optimisation.

Mofeed Turky Rashid received his BEng in Electrical Engineering from the University of Basrah, Iraq, in 1998, MSc in Electrical Engineering from University of Basrah, Iraq, in 2001, and PhD in Control and Computer Engineering from University of Basrah, Iraq, in 2011. He studied at the DIEEI Lab., Catania University, Italy, in the six-month fellowship program. He is currently a Professor at the University of Basrah, Iraq. He has held a Lecturer positions at several electrical and computer departments in Libya and Iraq for several years. He has many published articles in several national and international journals, in which his research interests cover the design and analysis of various control systems, DCS, machine learning, energy management and robotics systems.

1 Introduction

In many rural areas, people suffer from energy services lack because the communities of these areas are sparse and far from the grid, which makes grid installation is a difficult challenge due to many reasons related to weather, terrain, or distances, therefore energy cost in these areas is significantly become high (Zhang et al., 2020; Fioriti et al., 2017). In this regard, the microgrid (MG) that is developed with renewable energy resources (RER), integrated metering devices, and information technologies can cover the energy shortages in these areas (Jiang and Fei, 2015; Albadi and El-Saadany, 2007). Further, MG offers a wide range of benefits including system reliability enhancement and domestic energy supply, but the uncertainty of RER systems prevents the possibility of giving up on the fuel-based power systems. Therefore, the presence of DGs in MG became an inescapable issue that is a solution for the problem of energy supplying uncertainty (Al Hadi et al., 2020). The operation of RER side by side with DGs in the MG requires a complicated management process for several reasons associated for example with the generation, cost, or time of use. The consideration of RERs as unreliable, uncertain, fluctuant, and non-dispatchable resources (Qiu et al., 2016) make the power delivery unstable and difficult to manage also, led to use controllable and schedulable energy resources such as DGs and energy storage systems (ESS) to improve the efficiency and reliability of the MG system (Venayagamoorthy et al., 2016; Tushar et al., 2016). The provision of diesel generator operating fuel is one of the harsh tasks that must be met to ensure the operation reliability of these generators in hybrid MG. The meaning of hybrid MG is the network that consists of conventional and renewable energy sources. A significant problem of DGS operating related to the transport of the fuel from source locations to the rural areas, where this operation is stressful, expensive, and has many intricacies associated with delivering obligations and energy tariff policy.

In such cases, resorting to the principle of take-or-pay is one of the appropriate solutions to avoid these problems and hence ensuring fuel availability in the foreseeable future and overcoming the problem of deficiencies in the energy supply during peak times. As a result of the take-or-pay conception, there will be additional fuel consumption operation occur at specified times, and this leads to increase hours of generator operation,

that is because the energy management system (EMS) is forced to run the DG at time slots no need to operate in, to consume the remaining amount of fuel to avoid its accumulation to the next day. To decrease the effect of this problem, the dispatching and scheduling operation at times of need must be applied to the DG.

The operation of DG should be organised with other energy resources in the MG to overcome the dynamics behaviour of the system such as the uncertainty of RER, load curve shape fluctuations, or energy price variation from time to time (Zhang et al., 2019). Therefore the EMS is used to control the power flow between the energy resources on one side and the end-user on another side (Bui et al., 2020). To design an efficient EMS, two problems should be overcome, the first is the hybrid energy systems do not have a fixed model due to the stochastic nature of some parameters in the system like load demand, energy cost, or output of RERs (KEMA Inc., 2014). While the second is the consumer-side information ambiguity, where most consumers do not wish to give or disclose their privacy, therefore implementing a dynamic system and consumption scheduling is a challenge and difficult to apply traditional optimisation methodologies to manage such stochastic systems.

To deal with these barriers and cope with the randomness of system variables, the reinforcement learning (RL) approach can be developed as a possible solution for the optimal operation of MGs (Kim et al., 2016). The RL allows EMS to learn the behaviours of the customers, RER output, electricity cost variation, and other variables along the time horizon and analyse the system activities to obtain optimal decisions without prior knowledge about the system.

In recent years, many studies have appeared that considering the improvement of the performance of EMS in terms of fuel consumption to enhance the economic feasibility of the hybrid MG system as follow:

In Thirugnanam et al. (2018), battery energy management system is proposed, it reduces hours of the DGs operation using multiple types of batteries, while in Anglani et al. (2017) is presented an optimised energy management system to control islanded MG of a remote temporary military, DGs can be dispatched by load control using demand response as in Clavier et al. (2015) which focus on using the virtual capabilities of the demand side to perform diesel generation optimisation by the economic dispatch (ED). An EMS in Zhu et al. (2014) shares the optimal power among generators-set depending on fuel cost minimisation. In Wang et al. (2015), a decentralised power dispatch system for managed operation of multiple MGs and a distribution system, this paper introduces a methodology to depict the interactions between the distribution network operator and network of MGs. The short-term scheduling project of multiple fuel cell power plants connected simultaneously to supply electric and thermal energy to the community had been presented in El-Sharkh et al. (2010), where hybrid mechanisms depend on evolutionary programming (EP) and hill-climbing (HC) techniques to find the optimal solutions.

From the foregoing, it became clear that the process of scheduling and dispatching the operation of generators as well as determining their commitment is critical in formulating policy of energy generation and distribution, hence this reflected on the process of supplying and consuming fuel within the specified timelines, so it is necessary to schedule and allocate the fuel optimally. It is several efforts made in this field. In Lee et al. (1992), a practical adaptive fuel allocation method has been presented, the researchers in this paper use pseudo fuel price as an optimisation variable, while in

Kumar et al. (1984), a long-term fuel allocation and scheduling tackled in the optimum point to various generating units by using mixed and shared fuels subjected to long-term constraints. Within the scope of generation scheduling, Tong and Shahidehpour (1990) proposed a methodology for short-term unit commitment based on the Lagrangian Relaxation approach, on the other side, several methodologies are used to solve the unit commitment problem based on of priority lists scheme (PLS) in Baldwin et al. (1959), dynamic programming (DP) in van Meeteren (1984) or mixed integer programming (MIP) in Dillon et al. (1978).

Acquiring precise and a priori statistical information for all energy resources and loads in an MG is not easy, it may restrict applying of the aforementioned methods. Therefore to exceed the problem of stochastic models building, there are alternative methods (intelligent) that do not require prior information of the system model as in Jiang and Fei (2015) that introduces an energy ecosystem, a price effective smart MG depend on smart hierarchical agents with dynamic response and RER management, the intelligent method used is RL (Q-learning), but it suffers from the curse of dimensionality because of continuity of input states of the system (very large), therefore, deep neural network (DNN) as approximator added to RL to overcome the complexity and inefficiency of model-based methods, this method is also used in Du and Li (2020), where a multi-MG (MMG) energy management technique is presented depend on DNN and model-free RL to minimise the demand-side peak-to-average ratio (PAR). This research influences consumer preferences as a result of PAR decreasing, instead, they can substitute the shortage by DG operation scheduling with fuel allocation.

In Foruzan et al. (2018), they propose a multi-agent model to study EMS in an MG, the provider and users are modelled as autonomous agents, each one can make its decisions to maximise their profit in the environment using RL. A short-term stochastic optimisation is used in Fioriti et al. (2017) for a rural PV-diesel-battery hybrid MG to minimise fuel consumption and CO_2 emission.

All the aforementioned works did not address the issue of fuel scheduling and allocating independently using intelligent control strategies in a way that affects the energy management process in the network, as well as not taking into account the limitation of power flow at point of common coupling (PCC) together with the take-or-pay problem besides the impact of these constraints on the behaviour of the EMS.

To address the issues rose in earlier studies, this work presents an intelligent EMS consisting of a utility grid, PV array, ESS, and DG that is put in MG within a rural area. The goal is to dispatch the DG operation and schedule fuel consumption to assure its availability while appropriately consuming the allocated amount over the entire time horizon without relying on the system model (model-free), using a hybrid method that combines RL represented by Q-learning and DNN as approximator creating deep-Q-network, our system captures the stochastic nature of PV, load, and energy cost (DQN). The main contributions of this work are summarised below:

- Fuel allocation and consumption along the time horizon depend on the take-or-pay conception is presented. It is proposed an intelligent EMS responsible for the fuel consumption delivered from the supplier by scheduling the DG operation at the optimal times in a way that guarantees reducing costs and avoiding fuel accumulation to the next day.

- The RL is used as a model-free algorithm to regulate RER, ESS, utility grid, and DG operation to assure continuous and low-cost energy delivery to customers considering the premise of PCC limitation.
- Extending ESS life by training the EMS to regulate the charging and discharging process, taking into account the number of charging and discharging times lead to battery degradation life cycle.
- The time interval of the case study that will be worked on is the summer season that extends for three months, where the training takes place for the first two months and the test for the last month (66% of data for training and 34% for testing).
- The proposed method performance is verified by comparing the results with those obtained using the model-based optimisation methodology; in this work, the benchmark optimisation method is improved particle swarm optimisation (IPSO), and then the convergence of DQN to IPSO is determined.

The rest of this paper is organised as follows: Section 2 explains the system structure and the methodologies of EMS. Section 3 introduces the mathematical modelling of the system and the problem formulation. In Section 4, the result is presented and discussed. The performance is evaluated in Section 5, finally, the conclusion is provided in Section 6.

2 The proposed system

2.1 System structure

Figure 1 shows the system structure, it is consists of an MG connected to the grid at PCC, the MG comprises ESS, PV array, fuel tank, DG, and residential loads. The MG has intelligent EMS to control the energy resources operation. The power moves between the MG energy resources and household load in one direction and with the grid in two directions, meaning that the energy can be traded off (buying and selling) between the grid and MG. ESS, DG, and PV array are owned by the government. The essential purpose of ESS is to minimise the energy cost in a normal situation and enhance the reliability of the system during emergencies.

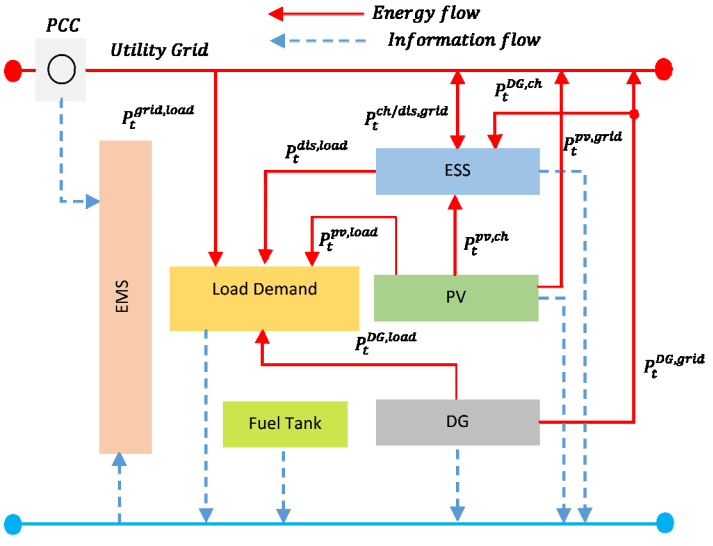
In peak times, ESS and DG can feed the power to the end-users to avoid the household load shedding and prevent their preferences from affecting. The controller tries to reduce the energy cost and consume the stored fuel in the tank during the day to avoid its accumulation to the next day as well as to avoid the penalty of non-use of the minimum fuel stipulated in the take-or-pay agreement. EMS makes the optimal actions and decides the amount of power to be traded with the grid taking into consideration the limitations of PCC and other constraints related to RER and fuel amount. The methodologies of the proposed work had been detailed in the next sections.

2.2 Intelligence of energy management system

As mentioned in the introduction section, the performance of EMS will be improved by the proposed RL technique to schedule and control the operation of MG, further,

optimising the consumption of DG fuel to reduce the operation cost. The proposed system is described as follow.

Figure 1 MG configuration (see online version for colours)



Note: The solid lines indicate power flow and the dashed lines indicate data flow.

2.2.1 Reinforcement learning

RL is a promising method, in which the most notable characteristic is model-free dependence, i.e., it is able to reach the optimal solution without depending on a priori knowledge or wearisome stochastic modelling (Oliehoek, 2012). The objective of RL is to find the optimal sequence of decisions (discrete or continuous) by exploring and exploiting the environment to be worked on without requiring prior knowledge of the system or predictive information (Busoniu et al., 2010). Within this scope, the agent interacts with the environment by making a series of actions (decisions) depending on the states (inputs) of the system and the value of the reward as a result of these actions. RL comprise the term Markov decision process (MDP), which consists of four-term shown below and its dynamics illustrated in Figure 2 (Dillon et al., 1978):

1 Reward function r: $\mathcal{S}\times\mathcal{A}\rightarrow\mathbb{R}$.

2 States vector $\mathcal{S}$: the state s_t at time slot t is the status of the environment.

3 Actions vector $\mathcal{A}$: action at means the decision that the agent takes at time slot t.

4 Transition dynamics with conditional transition probability $p(s_{t+1}|s_t, a_t)$ satisfying the Markov property, i.e.,

$$p\left(s_{t+1}\middle|s_t, a_t\right) = \Pr\left(s_{t+1}\middle|s_1, a_1, \ldots, s_t, a_t\right) \tag{1}$$

2.2.2 Q-learning

One of the most important penetrations in RL was the development of Q-learning. It is a standard form of one-step Q-learning is defined by the Bellman equation as shown in equation (2) (Sutton and Barto, 2014).

$$\underset{s \leftarrow s_{t+1}}{Q}(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \cdot \left[r_{t+1} + \gamma \cdot \max_a Q(s_{t+1}, a_t) - Q(s_t - a_t) \right] \quad \forall t \in T \qquad (2)$$

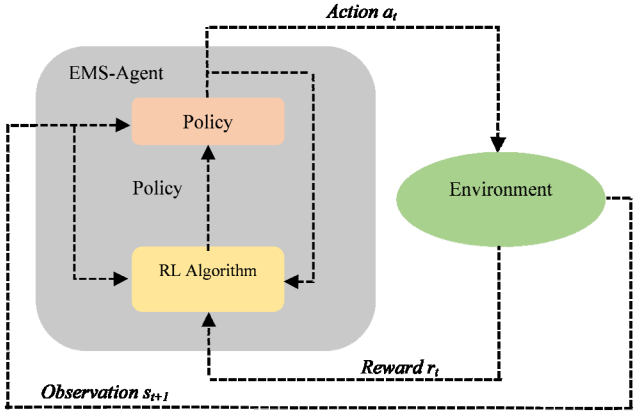
where $Q(s, a)$ is Q-value of current state s and action a pair, α is learning rate, γ is discount factor of future rewards, and r is immediate reward value.

The agent in Q-learning explores the environment in order to know the states and take actions to obtain stacked rewards and then move to another state and so on (Bui et al., 2020; Sutton and Barto, 2018). Generally, for economic dispatching in the MG, the state vector should be continuous. As a result, the conventional Q-learning technique cannot interact with this issue because it suffers from the curse of dimensionality problem, where it depends on a look-up table to represent the Q-value function for each state-action pair. To overcome this problem, a DQN had been used, which exploits the approximation properties of DNN (Liang and Srikant, 2017; Huang et al., 2020).

2.2.3 Deep Q-learning method

In Q-learning, a Q-table work on discrete state space, therefore, in continuous state space this method become more intractable (Bui et al., 2019), so DNN comes to play the role of approximator, which can take state inputs as a vector and learn to map them into Q-values for all possible actions, this operation called DQN.

Figure 2 The dynamics of EMS-agent training (see online version for colours)



2.3 Improved particle swarm optimisation

The particle velocities in PSO build up rapidly and this may lead to skipping the maximum of the objective function. To solve this problem IPSO is suggested by adding an inertia term w to decrease the velocity and enable the particle to converge more

precisely and efficiently compared with PSO. Usually, the value of w is ranged between 0.4 and 0.9 as the iteration progresses. The j^{th} particle velocity becomes as shown:

$$V_j(i) = wV_j(i-1) + c_1 r_1 \left[P_{best,j} - x_j(i-1)\right] + c_2 r_2 \left[G_{best} - x_j(i-1)\right]; \quad j = 1, 2, \ldots, N \quad (3)$$

Equation (3) refers to the velocity of particle j in the i^{th} iteration. The coefficients c_1 and c_2 are the individuals and social learning rates respectively, and r_1 and r_2 are distributed random numbers between 0 and 1.

N represents the number of particles, $P_{best,j}$ is the historical best value of individual $x_j(i)$ that has the highest objective function value in the current iteration, while G_{best} is the historical best value of all particles until the current iteration that has the highest objective function value in all the previous iterations.

Equation (3) shows a larger value of w support the global point of optimisation and a smaller value support a local search. Thus a large value of w makes PSO explore new regions without many local optima and this consequently leads to failure in finding the true optimum. A solution to this problem a balance must be achieved between global and local exploration, this done by making inertia value decreases linearly with the iteration number as illustrated in equation (4) (Tushar et al., 2016):

$$w(i) = w_{\max} - \left(\frac{w_{\max} - w_{\min}}{i_{\max}}\right) i \quad (4)$$

where $w_{\max}$ and $w_{\min}$ are maximum and minimum value inertia weights respectively (recommended values are $w_{\max} = 0.9$, $w_{\min} = 0.4$), $i_{\max}$ is the maximum iteration number used in the algorithm.

3 The mathematical model of the proposed system and problem formulation

3.1 The mathematical model

The MG is connected to the utility grid in PCC. The energy move across PCC is specified by a limited amount of kw/h that cannot be exceeded due to the difficulty of delivering energy to rural communities in sufficient quantity. The conceptual architecture of the proposed hybrid MG depicted in Figure 1 is modelled as follows.

3.1.1 PV model

The PV power output at time slot t is obtained depending on the irradiance intensity, PV cell temperature T_c and ambient temperature as illustrated in (5) (Michaelson et al., 2018):

$$T_c = T_t + \left(\frac{G_t}{800}\right)(T_t - 20) \qquad \forall t \in T \quad (5)$$

where T_{noc} is the nominal operating cell temperature. The PV power can then be calculated as given in (6):

$$P_t^{PV} = \left(G_t / G_{stc}\right) P_{\max}^{PV} \left(1 + k_c \left(T_c - T_{stc}\right)\right) \eta_{MPP} \qquad \forall t \in T \quad (6)$$

where $G(t)$ is forecast irradiance, G_{stc} is standard test condition irradiance, k_c is relative temperature coefficient.

3.1.2 Battery model

The battery usage aims to store extra power of the PV and that imported from the grid to utilise it when there is a shortage in meeting up the demand or it is below the minimum starting threshold of DG. The ESS used in this paper is modelled by equation (7) (Elkazaz et al., 2020; Mandal et al., 2018):

$$E_t = E_t + \Delta t * P_t^{ch} * \eta^{ch} * u_t^{ch} - \Delta T \left(P_t^{dis} / \eta^{dis} \right) * u_t^{dis} \qquad \forall t \in T \tag{7}$$

where u_t^{ch} and u_t^{dis} are logic variables to control battery charging/discharging, as derived in equations (8) and (9).

$$u_t^{ch} = -action_t^1 * \left(1 - action_t^2\right)/2 \qquad \forall t \in T \tag{8}$$

$$u_t^{dis} = action_t^1 * \left(1 + action_t^1\right)/2 \qquad \forall t \in T \tag{9}$$

The state of charge (SOC) is considered one of the important factors in the ESS management, it represents the ratio of energy in the ESS bank and is modelled as in equation (10).

$$SOC = E_t / E_{bat,\max} \qquad \forall t \in T \tag{10}$$

3.1.3 Diesel generator model

The DG is installed in the MG as dispatchable energy resources to ensure the reliability of energy delivery, it can offer a flexible backup source to supply the power at peak intervals or at shortage times in the MG.

The fuel amount consumption per hour of the diesel generator depends on the dispatched power drawn from it and on its rated power and it is modelled by the linear equation as shown in equation (11):

$$F_t = \alpha_{DG} * P_t^{DG} + \beta_{DG} * P_t^{DG,rated} \qquad \forall t \in T \tag{11}$$

where α_{DG} and β_{DG} are coefficients that represent incline and interception of the fuel consumption curve from the manufacturer with typical values equal 0.246 l/kWh and 0.815 l/kWh respectively (Kaabeche and Ibtiouen, 2014; Ismail et al., 2013; Wu et al., 2016).

3.2 Objective function and problem formulation using RL

3.2.1 Objective function

Two main objectives must be achieved in the MG, the first is to minimise the cost of the supplied energy at each time slot by increasing the dependence on RER, the second is to optimise the consumption of entire daily fuel, which is supplied to the DG according to the terms of the signed agreement between the supplier and the operator in the MG relying on the take-or-pay conception. For the two objectives at each time slot t in the

horizon, the decision variables (actions) are to be battery charging/discharging and DG operation dispatching. The EMS uses DQN to optimise the operation of resources in order to meet net demand and fuel consumption until the end of each day.

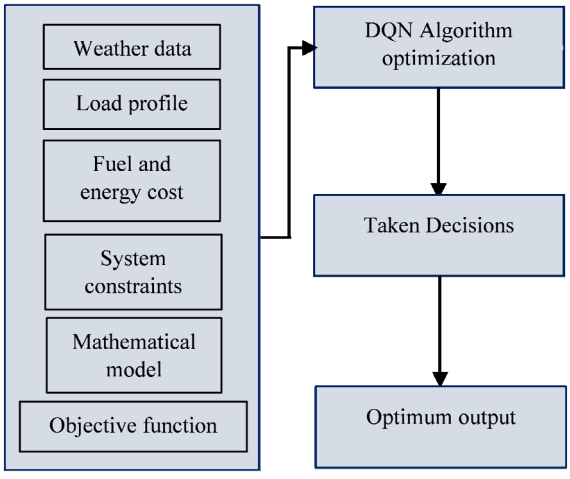
A framework of optimisation technique has been introduced to achieve EMS objectives. Figure 3 illustrates the block diagram of the framework (Mandal et al., 2018), which passes through four phases to get the output as shown in the internal blocks. At each time slot, the operation cost includes three parts: imported energy cost from the grid, DG energy cost, and battery degradation cost due to the charging and discharging process. The formulation of the energy cost minimisation is represented in equation (12).

$$\min \sum\nolimits_{t\in T} C_t^g\left(P_t^g\right)+C^{DG}\left(P_t^{DG}\right)+C^{\deg} \qquad \forall t\in T \tag{12}$$

$$C^{DG}=\left(F*C^f\right)/P_t^{DG} \qquad \forall t\in T \tag{13}$$

$$C^{\deg}=\rho^{\deg}\left(\left|SOC_t-SOC_{t-1}\right|\right) \qquad \forall t\in T \tag{14}$$

Figure 3 The framework of the optimisation



The first term in the objective function (12) is the cost of purchased energy from utility grid at time slot t where the retail price of energy changes from time to time at PCC, the second term is the DG generation cost at time slot t, the price in DG depend on fuel and maintenance cost as illustrated in equation (13), it also depends on the level of generation, as the more generation the lower price/kWh, the maintenance cost can be neglected because it a small compared with the hourly cost of energy generation, the third term is the battery degradation cost due to frequent charging/discharging as shown in equation (14) along time horizon where the change between two consecutive battery actions lead to battery life degradation (Liu et al., 2018; Aklo and Rashid, 2021).

3.2.2 Problem formulation using RL

An MDP is applied to formulate the energy management dilemma in Figure 1. EMS contains the agent that tries to learn how to make optimum actions that represent the best

scheduling of energy resources operation and fuel consumption strategy, this is done by repeated interactions with the environment. In this environment, the transitions of the states from t to $t + 1$ in state space are depending on states values at time slot t and uncertainty of the energy demand, price, and PV power generation.

In this work, RL formulation is divided into three fundamental parts:

1 a set of the system (environment) states $\mathcal{S}$

2 a set of controller (agent or EMS) actions $\mathcal{A}$

3 a sequence of rewards rt at each time slot along the entire time horizon.

In the proposed system the state space length is a six-dimension vector:

$$s_t = \left[t, SOC, SOF, P_{pv}, P_L, Day\right] \in \mathcal{S} \tag{15}$$

where t denotes time slot index (one hour in this work), SOC is the state of charge in ESS, it is measured as a percentage ratio, state of fuel (SOF) is the state of fuel, where its value indicates the fuel amount in the tank. P_L is load demand and *Day* represents the day rank in the month, P_{pv} is output generation of the solar. The actions are two-dimension vector, *action*1 for battery charging/discharging control, and has three cases while *action*2 to control DG operation in seven levels as shown in equations (16) and (17).

$$action1 = \left\{a_t^{ch}, a_t^{dis}, a_t^{idle}\right\} \in \mathcal{A} \tag{16}$$

$$action1 = \left\{a_t^{11}, a_t^{12}, a_t^{13}, a_t^{14}, a_t^{15}, a_t^{16}, a_t^{17}\right\} \in \mathcal{A} \tag{17}$$

where the cases a_t^{ch}, a_t^{dis}, $a_t^{idle} \in \{-1, 1, 0\}$ belong to the battery action, where '–1' for charging, '1' for discharging, and '0' for idle state. The other action consists of seven cases: a_t^{11}, a_t^{12}, a_t^{13}, a_t^{14}, a_t^{15}, a_t^{16}, $a_t^{17} \in \{0, 0.35, 0.5, 0.6, 0.75, 0.9, 1\}$, each one of these cases represent a single level of the rated power delivered from DG, where there are seven working levels in the DG, they are 0%, 35%, 50%, 60%,75%, 90% and 100% of rated power respectively. The reward function is derived by considering the following:

1 energy cost-minimising

2 diesel generator operation dispatching

3 fuel consumption according to take-or-pay provision

4 ESS charging and discharging optimisation (degradation reduction).

All these are done within the system constraints such as PCC constraint, fuel storage constraints. In this regard, RL trains the agent to reach the highest value of the reward by trying a sequence of actions to find an optimal policy that leads to the lowest energy cost. It actually performs a comparison between energy prices for both sources in addition to required energy at each time slot taking into account PV power which is considered very cheap or free compared with other resources (Du and Li, 2020).

3.2.3 System constraints

The constraints of EMS comprise power balance constraints, energy capacity constraints, and operational constraints. The power balance constraint states the total generated power

from different kinds of energy resources should be equal to the load demand in the MG. The power balance between generation and consumption is depicted by equation (18).

$$\sum_{t\in T} P_t^g + P_t^{PV} + P_t^{DG} + P_t^{dis} = \sum_{t\in T} P_t^L + P_t^{ch} \tag{18}$$

where P_t^g is trading off power between the MG and utility grid, when it is positive, this means buying power from the grid and selling when it is negative and zero means no power exchange is between, while P_t^L is the load demand.

The constraint associated with the grid is included in equation (19), it shows the upper limit of purchasing power that cannot be exceeded in PCC.

$$P_t^g \le P^{PCC} \qquad \forall t \in T \tag{19}$$

The constraints of the DG are given by equations (20) and (21). Equation (20) shows the maximum and minimum bounds of dispatched power output from DG, where k is set to be 0.35 according to the recommendation of manufacturers. Equation (21) is an on/off status of DG.

$$z_t * k * P^{DG,rated} \le P_t^{DG} \le z_t * P^{DG,rated} \tag{20}$$

$$z_t = \begin{cases} 1 & \text{DG is on} \\ 0 & \text{DG is off} \end{cases} \qquad \forall t \in T \tag{21}$$

ESS constraints are listed in equations (22), (23), (24), (25) and (26), equations (22) and (23) show the upper and lower limits of the battery rated power in both charging and discharging, equation (24) illustrates the allowable percentage ratio of the battery capacity limits while equation (25) is the constraint of control variables in order to prevent charging/discharging at the same time, its value is '0' or '1' as it is seen in equation (26).

$$0 \le P_t^{ch} \le u_t^{ch} * P_t^{ch,\max} \qquad \forall t \in T \tag{22}$$

$$0 \le P_t^{dis} \le u_t^{dis} * P_t^{dis,\max} \qquad \forall t \in T \tag{23}$$

$$SOC^{\min} \le SOC_t \le SOC^{\max} \qquad \forall t \in T \tag{24}$$

$$0 \le u_t^{ch} + u_t^{dis} \le 1 \qquad \forall t \in T \tag{25}$$

$$u_t^{ch}, u_t^{dis} \in \{0, 1\} \qquad \forall t \in T \tag{26}$$

Boundaries of SOF are illustrated in equations (27), (28), and (29), where SOF is the ratio of the current capacity of fuel in the tank to the total capacity of the tank.

$$SOF_t = SOF_{t-1} \mp F_t / F^{\max} \qquad \forall t \in T \tag{27}$$

$$SOF^{\min} \le SOF_t \le SOF^{\max} \qquad \forall t \in T \tag{28}$$

The take-or-pay provision penalising the buyer for not purchasing the minimum quantity of fuel within a specified period is represented by equation (29):

$$SOF_{T/P} = SOF^{\min} \qquad \forall t \in T \tag{29}$$

where $SOF^{\min}$ is a percentage that represents the amount of fuel remaining in the tank at the end of the day that should be available for emergency cases.

Remark 1: In order for the EMS to work perfectly, the parameter values must be taken in a manner that simulates the actual reality and corresponds to the real power grid and industrial devices. Therefore, the load demand, PV generation data, the utility grid energy cost, the DG and ESS parameters had been taken from realistic scenarios.

Remark 2: The imperfections are an unavoidable issue in energy generation processes, and even though power plants run far from ideal conditions, they still work. This is because imperfections lead to rising hidden dynamics, which, when triggered at the right time, have a beneficial overall effect on these devices. Imperfection is can stem from the performance of energy resources or even from the procedures of their generation. The traditional approach toward imperfection is to think of it as a source of uncertainty. Controlling imperfect systems is particularly desirable when a big number of units are involved, resulting in a system with a large number of state variables (Bucolo et al., 2019), but in this work, this issue is treated depending on techniques of DNN, and this subject was left to the future work.

4 Simulation results and analysis

Simulations had been performed on SM consisting of central PV array, DG, and ESS. The DG-rated power calculation depends on the historical highest load demand. ESS is connected to PCC, DG, and PV on one side and to the households loads on the other side as shown in Figure 1, ESS can be charged from the surplus power of PV or the utility grid at optimal times according to the price of energy at a current time slot or according to the load demand, all of these activities depend on balance power equation (18) and the optimisation policy.

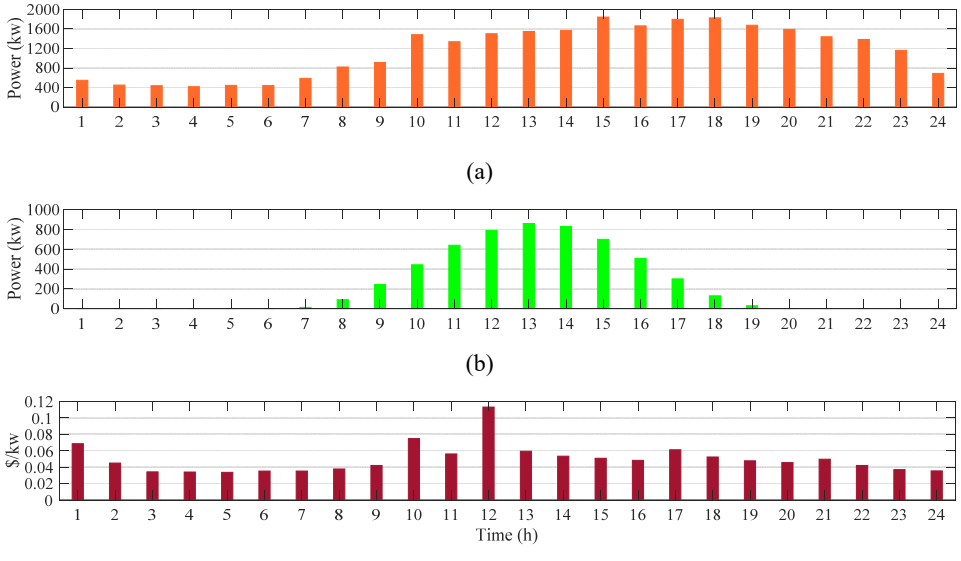
DQN is applied on real data taken from the independent electricity system operator centre (available on https://www.ieso.ca), which works at the centre of Ontario's power system. The proposed work is applied on 200 households, one-hour time step, 24-hour rolling horizon, and training interval are on summer season in Ontario city that start from 21 June until 21 September, the historical data of the first two months (21 July to 21 August) will be used to train the agent and the last month data (until 21 September) for testing where the three months have the same characteristics in terms of weather and other factors. The electricity buying price depends on the time-of-use tariff concept. Table 1 illustrates the details of the work.

After the agent training is completed, the optimal policy becomes ready to deploy on the model to determine the decisions in real-time. To validate the algorithm efficiency, the agent policy had been applied on a typical day elected from training data and on another typical day from testing data. Figure 4 illustrates the load demand, PV power generation, and energy cost profiles of the typical day respectively within training data. Figure 5(a) shows the two actions that were made as a result of deploying the optimum agent policy on the training day.

Table 1 Parameters of MG resources

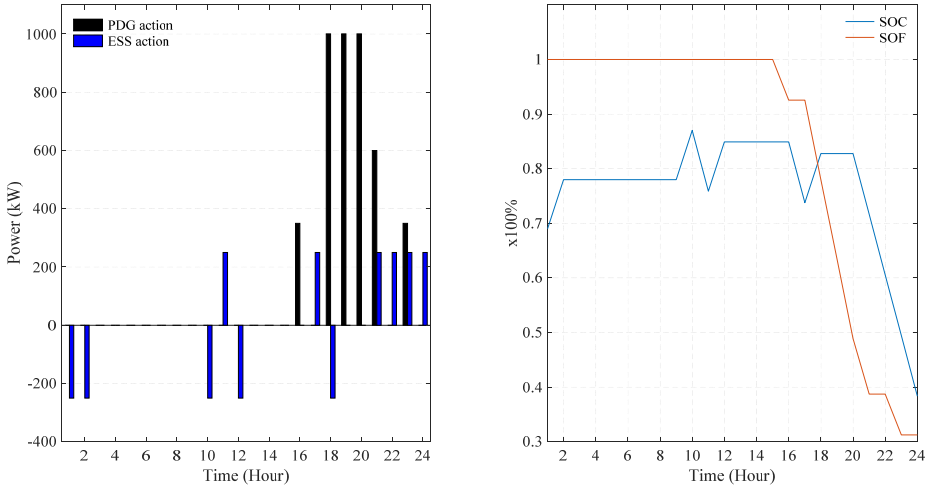
CBESS	SOC^{min} (%)	SOC^{max} (%)	$P_t^{ch,max}$ (kWh)	$P_t^{di,smax}$ (kWh)
	20	90	250	250
	E^{max} (kW)	η^{ch} (%)	η^{dis} (%)	ρ^{deg} ($/MW)
	2,500	0.9	0.9	100
DG	$P_t^{DG,min}$ (kWh)	$P_t^{DG,rated}$ (kWh)	α_{DG} (l/kWh)	β_{DG} (l/kW)
	350	1,000	0.246	0.08145
	F^{max} (litre)	C^f ($/litre)	SOF_{min} (%)	
	2,250	0.75	30	
PV	P_{pv}^{max} (w)	T_{noc} (CO)	G_{stc} (W/m²)	T_{stc} (C^O)
	200	44.5	1,000	25
	k_c (%/Co)	η_{MPP} (%)		
	0.43	90		
PCC	$P^{PCC,max}$ (MW)			
	1.5			

Figure 4 Typical training day data, (a) load demand profile (b) PV generation profile (c) grid energy cost profile (see online version for colours)



It shows charging/discharging action in addition to DG commitment and scheduling action. The results indicate that the training using the proposed algorithm gives good results in terms of commitment and energy cost reduction as will be proved. The main two constraints in the EMS are the fuel consuming ratio and capacity limits of PCC along the entire time horizon.

Figure 5 (a) Optimum agent actions of the training day (b) SOF in the tank and SOC in the ESS of the training day (see online version for colours)



(a) (b)

First, EMS responsibility within this scope is to learn in what time step within the time horizon the DG can be run and at what level of operation to obtain energy at the lowest cost, while ensuring all the supplied fuel for the current day is spent according to the take-or-pay conception to avoid any penalties. According to the made decisions, the status of SOC and state of SOF became as shown in Figure 5(b).

Table 2 SOC of the training day along the time horizon

Hour	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
SOC (%)	69	78	78	78	78	78	78	78	78	87	75	84
Hour	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
SOC (%)	84	84	84	84	73	82	82	82	71	60	49	38

Table 3 SOF of the training day along the time horizon

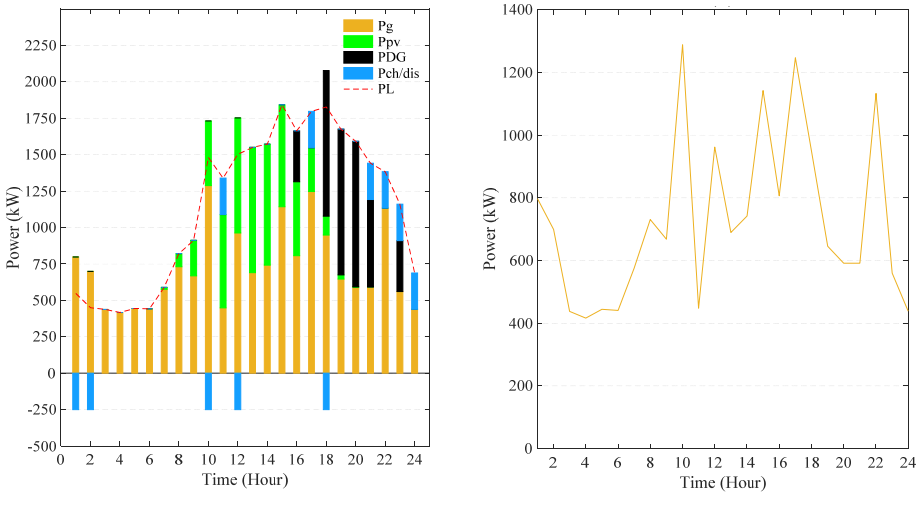
Hour	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
SOF (%)	100	100	100	100	100	100	100	100	100	100	100	100
Hour	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
SOF (%)	100	100	100	92	92	78	63	48	38	38	31	31

It is noted from the ratio curves that the agent does not allow to exceed the boundaries of the constraints for these parameters as illustrated in Table 2, where the SOC along the time horizon did not exceed 90% of the maximum capacity and was not less than 20% and at the end of the day. In the related context, it is noticed that ESS is charged at the beginning of the day because the demand is light and the charging energy cost is low, also it is charged at time slots when there is a surplus of the energy generated by PV (hours 10, 12) as shown in Figure 5(a). In contrast, the discharge happens at peak times when the demand is large and the energy supplied from the grid is expensive or shortages

(hours 11, 17, 21–24). In a related context, SOF became exactly 30% at the end of the time horizon as specified in Table 1, where, the agent is committed to consuming the entire amount of allocated fuel until 31% of fuel capacity for the current day as shown in Table 3, where the error ratio between the actual and desired ratio is equal to 1%.

It is clear from Figure 6(a) that detailed in Table 4 the proposed algorithm learns the control system to obtain the most benefit of energy cost decreasing by exploiting the energy generated from PV and ESS to cover a significant part of the required energy during peak times where PV reduce the peak of the daytime (hours 8–17) while ESS reduce the peak of the night (hours 17–23). The end-user needs to purchase the rest of his required energy from the grid at times of no discharging, or no PV generation or may few when the weather is cloudy or rainy, the agent import the power from the utility grid within the boundary of PCC (in this research is about 1.5 MW) But in the event of insufficient generation from all these sources, the agent resorts to operating a diesel generator to fill the generation shortfall, this done usually at peak times (hours 16, 18–21, and 23) as shown in Figure 6(a). Figure 6(b) shows the amount of energy imported through PCC, where it is observed that the maximum power is 1.288 MW at hour 10 and does not exceed the boundaries of the PCC, this shows how well the control system adheres to the parameters constraints was trained.

Figure 6 (a) The distribution of the energy in the training day (b) Energy imported from the grid (see online version for colours)



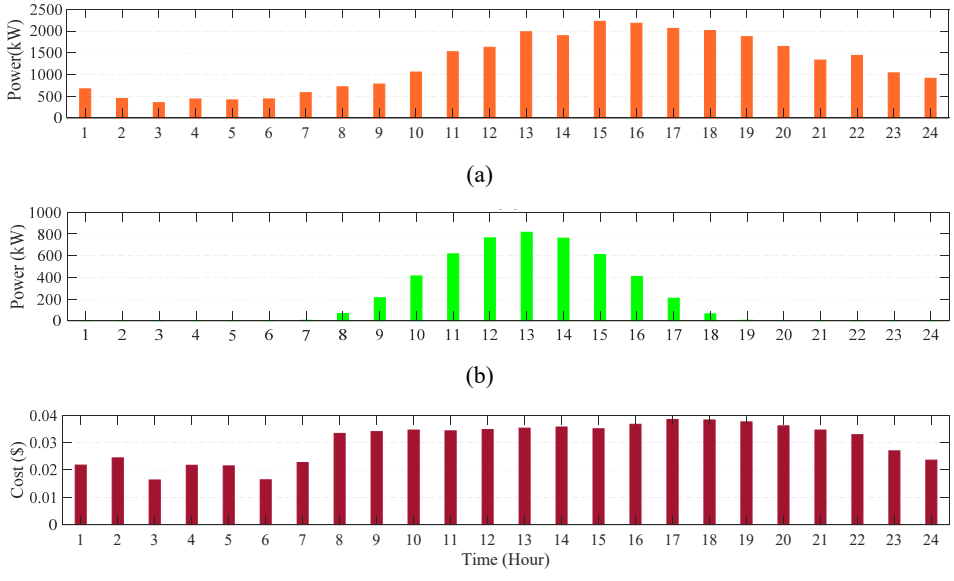
(a) (b)

Despite one of the tasks of the EMS is to cover the energy demand, there is another task that is getting rid of the surplus fuel by operating the generator at times of no actual need. It is observed that the energy generated from the DG was used to cover the demand without resorting to purchasing energy from the grid to avoid losses resulting from the accumulation of fuel to the next day according to the take-or-pay agreement.

Table 4 Power distribution and balancing of the training day (kW)

Hour	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
PL	548	449	438	417	444	441	589	821	912	1,481	1,339	1,503
Pg	798	699	438	417	444	441	575	731	668	1,288	448	962
Ppv	0	0	0	0	0	0	13	90	244	443	641	790
PDG	0	0	0	0	0	0	0	0	0	0	0	0
ESS	–250	–250	0	0	0	0	0	0	0	–250	250	–250
Hour	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
PL	1,550	1,572	1,842	1,664	1,797	1,827	1,675	1,592	1,441	1,383	1,160	688
Pg	689	742	1,142	806	1,247	948	645	592	591	1,133	560	438
Ppv	860	830	700	507	301	129	30	0	0	0	0	0
PDG	0	0	0	350	0	1,000	1,000	1,000	600	0	350	0
ESS	0	0	0	0	250	–250	0	0	250	250	250	250

Figure 7 Typical testing day data, (a) load demand profile (b) PV generation usage (c) grid energy cost profile (see online version for colours)

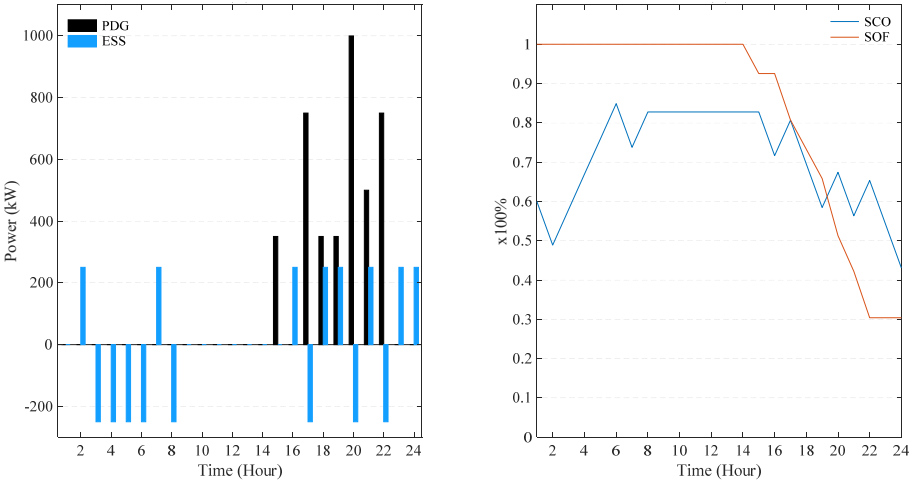


To validate the training process and the accuracy of the learning results, the policy was tested on a typical day to be chosen from the last month of the summer season (21 August to 21 September). Figure 7 illustrates the load demand, PV power generation, and energy cost profiles for the testing day, respectively. Comparing these figures' profiles with that of the previous training day, it is noted that they have the identical tendency. It is seen from load profiles, the peak and valley times are near for the two days, as well as the generation ratio for PV array close to a large extent, as is the case

for the energy tariff price, this leads to making the result of testing is close to training results because of adopting the deployed policy of the training, see Figures 5(a) and 8(a).

It is noted that the charging and discharging decisions in the two days is almost identical, where the charging is done in times of light load and the low price of energy tariff and the discharge is at peak times and the high price of energy, also, in the energy deficiency intervals, the decisions of DG operating at different levels are made in the same region approximately of the time horizon of the two days. From Figure 8(b), it is observed that there is a clear commitment to the constraints imposed on the system as shown in Tables 5 and 6, where the charging ratio does not exceed 90% and not less than 20%, also, the other system constraint related to the use of the entire amount of fuel until to 30% of tank capacity.

Figure 8 (a) Optimum agent actions of the testing day (b) SOF in the tank and SOC in the ESS of the testing day (see online version for colours)



(a) (b)

Table 5 SOC of the testing day along the time horizon

Hour	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
SOC (%)	60	48	57	66	75	84	73	82	82	82	82	82
Hour	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
SOC (%)	82	82	82	71	80	69	58	67	56	65	54	43

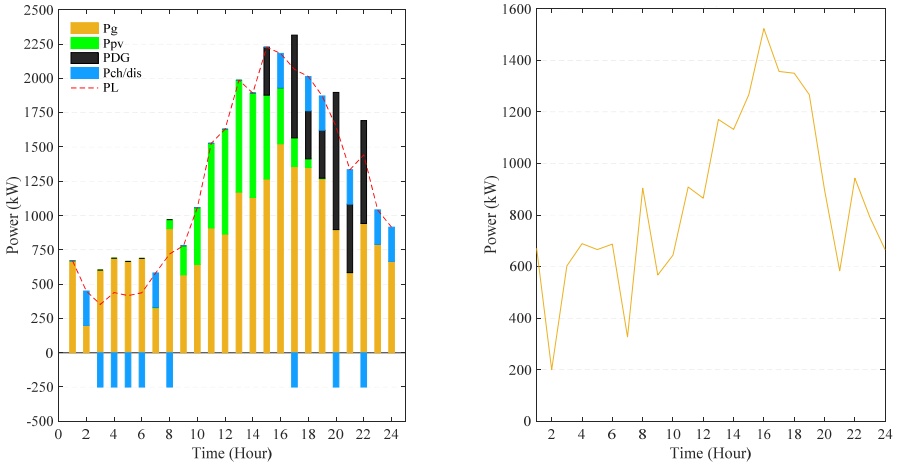
Table 6 SOF of the testing day along the time horizon

Hour	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
SOF (%)	100	100	100	100	100	100	100	100	100	100	100	100
Hour	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
SOF (%)	100	100	92	92	80	73	65	51	42	30	30	30

In addition to other constraints such as that related to the minimum operation of the generator as in Figure 8(a) or PCC limits keeping as it is evident in Figure 9 that shows

also the distribution of generated power of each energy resource at each time slot according to the balanced power equation (18).

Figure 9 (a) The distribution of energy in the testing day (b) Energy purchased from the grid in the testing day



(a) (b)

It is observed from Figure 9(b) that the energy demand is covered by the same way in Figure 6, where the solar energy is fully utilised to cover the demand at daytime (hours 9–16) and charging the battery (hours 12, and 18), while the generator is run at peak times to decrease the grid burden. Table 7 illustrates the details.

Table 7 Power distribution and balancing of the testing day (kW)

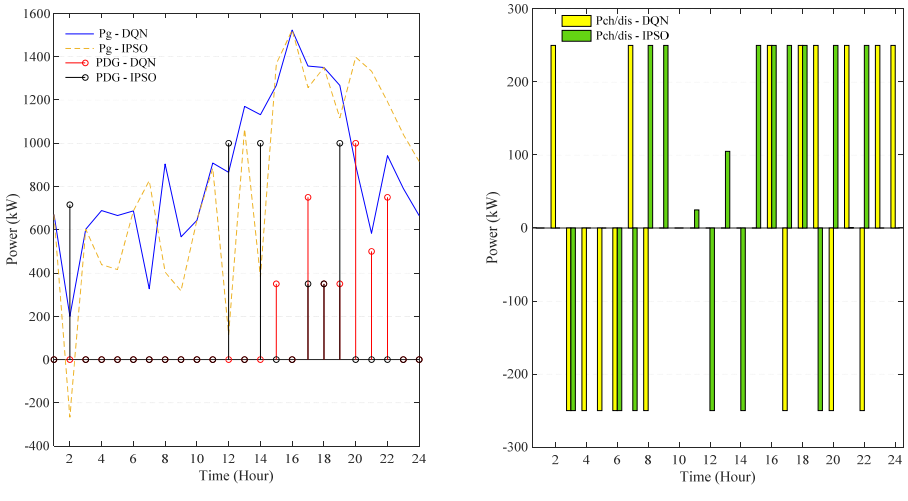
Hour	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
PL	671	450	352	439	416	437	582	720	780	1,057	1,526	1,629
Pg	671	200	602	689	666	687	327	904	567	644	909	865
Ppv	0	0	0	0	0	0	5	66	213	413	617	764
PDG	0	0	0	0	0	0	0	0	0	0	0	0
ESS	0	250	–250	–250	–25	–25	250	–250	0	0	0	0
Hour	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
PL	1,986	1,892	2,227	2,181	2,066	2,013	1,871	1,648	1,334	1,443	1,041	915
Pg	1,171	1,132	1,267	1,523	1,357	1,350	1,267	898	584	942	791	665
Ppv	816	761	611	408	209	63	5	0	0	0	0	0
PDG	0	0	350	0	750	350	350	1,000	500	750	0	0
ESS	0	0	0	250	–250	250	250	–250	250	–250	250	250

5 Performance evaluation

To benchmark DQN algorithm performance, IPSO method has been used as a model-based controller to minimise the daily energy demand cost. For IPSO, assuming full knowledge of the system parameters and perfect estimation of the uncertain parameters such as energy cost, PV power, or load demand. Consequently, the results obtained from this method are certain and considered a criterion for measuring the quality of the results in DQN.

The performance of the proposed method is investigated by IPSO in terms of two factors, energy generation factor, and energy cost factor. Figure 10 shows the energy generation from the grid, DG, and charging/discharging of ESS using DQN and IPSO. To prove the efficiency, the amount of imported power from the utility grid and DG generation are compared in the two methods then the error ratio is calculated. Table 8 shows the difference of grid energy between the two methods at each time slot and the error ratio is computed, it is equal to 4% as shown at the end of Table 8, this ratio reflects the robustness of this method that does not need the system modelling for energy management.

Figure 10 Energy generation of the testing day using DQN and IPSO, (a) energy generation of grid and DG (b) energy charging/discharging (see online version for colours)



(a) (b)

On the other hand, Table 9 indicates an error ratio reaches 9% of energy generation using DG, but this ratio has a small effect on the overall energy consumption because the contribution of the DG in energy supplying to cover the load demand is small compared with the total energy generation of other energy resources.

In term of energy cost, the cost of the supplied energy is divided into three parts, the first part is the cost of energy imported from the utility grid, the second part is relevant to the cost of the energy generated from DG which includes the cost of fuel, maintenance, and operating, and the last part is ESS degradation cost.

Table 8 Utility grid energy statistic using DQN and IPSO in the testing day (kW)

Hour	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
Pg-DQN	671	200	602	689	666	687	327	904	567	644	909	865	1,171
Pg-IPSO	671	–266	602	483	416	687	827	404	317	644	884	115	1,066
Difference	0	400	0	206	250	0	500	500	250	0	25	750	105
Hour	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*Total*	*Error*
Pg-DQN	1,132	1,267	1,523	1,357	1,350	1,267	897	584	943	791	665	20,678	4%
Pg-IPSO	382	1,367	1,523	1,257	1,350	1,117	1,398	1,333	1,193	1,041	915	19,726	
Difference	750	100	0	100	0	150	501	749	250	250	250	952	

Table 9 DG statistic using DQN and IPSO in the testing day (kW)

Hour	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
Pg-DQN	0	0	0	0	0	0	0	0	0	0	0	0	0
Pg-IPSO	0	716	0	0	0	0	0	0	0	0	0	1,000	0
Difference	0	716	0	0	0	0	0	0	0	0	0	1,000	0
Hour	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*	*Total*	*Error*
Pg-DQN	0	350	0	750	350	350	1,000	500	750	0	0	4,050	9%
Pg-IPSO	1,000	0	0	350	350	1,000	0	0	0	0	0	4,416	
Difference	1,000	350	350	400	0	650	1,000	500	750	0	0	366	

Figure 11 demonstrates the cost of bought energy from the utility grid using DQN and IPSO algorithms. In Figure 11, the hourly cost of the energy imported from the utility grid using IPSO and DQN algorithms is illustrated, where it is noted that the cost is neighbouring using the two methods and the average cost per hour is \$26.25/h using IPSO and \$27.7/h using the DQN and the error ratio between the two values is 5% as shown in Table 10. The same for the cost of DG energy (Figure 12) and the ESS degradation cost (Figure 13), where the average cost for the two algorithms is very close and the error ratio is 0.5% and 8% as is evident in Table 10.

Figure 11 The hourly cost of grid energy using IPSO and DQN (see online version for colours)

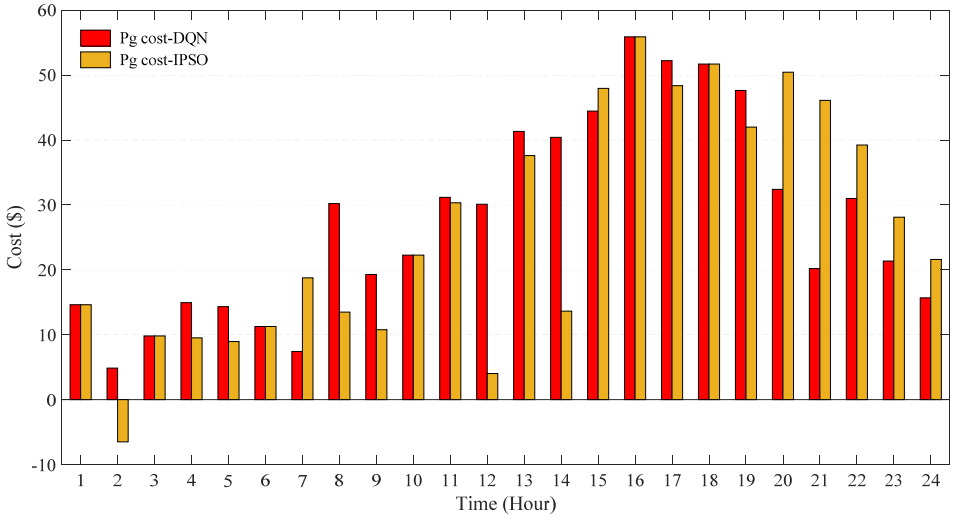


Figure 12 The hourly cost of generator energy using IPSO and DQN (see online version for colours)

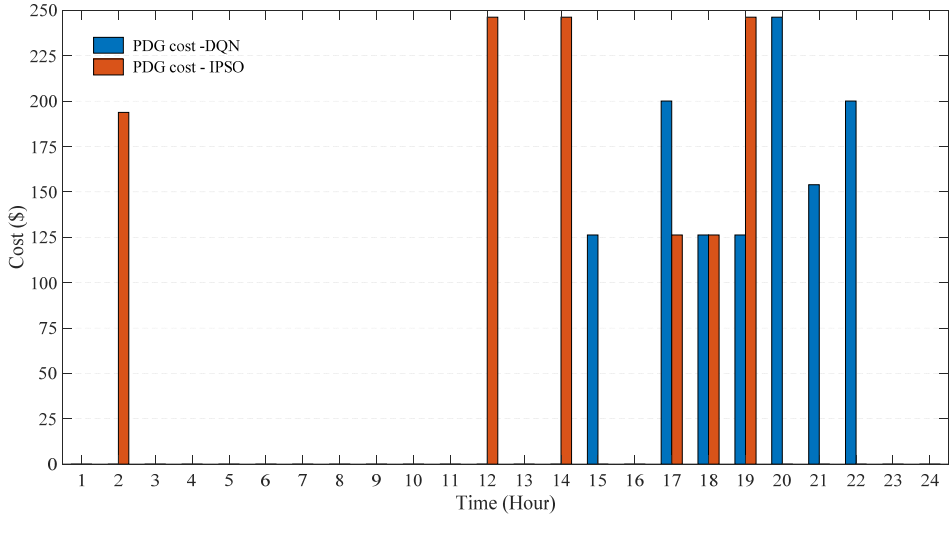


Figure 13 The hourly cost of battery degradation using IPSO and DQN (see online version for colours)

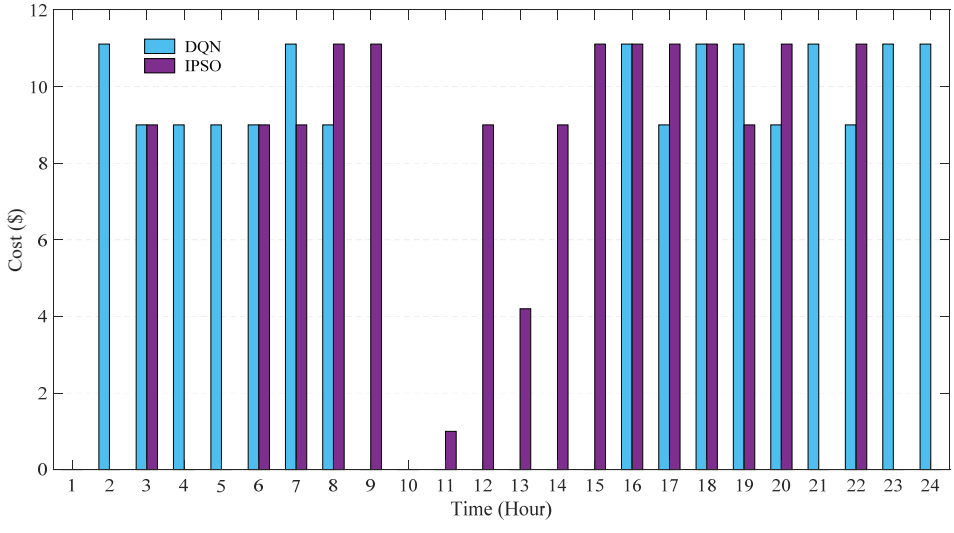
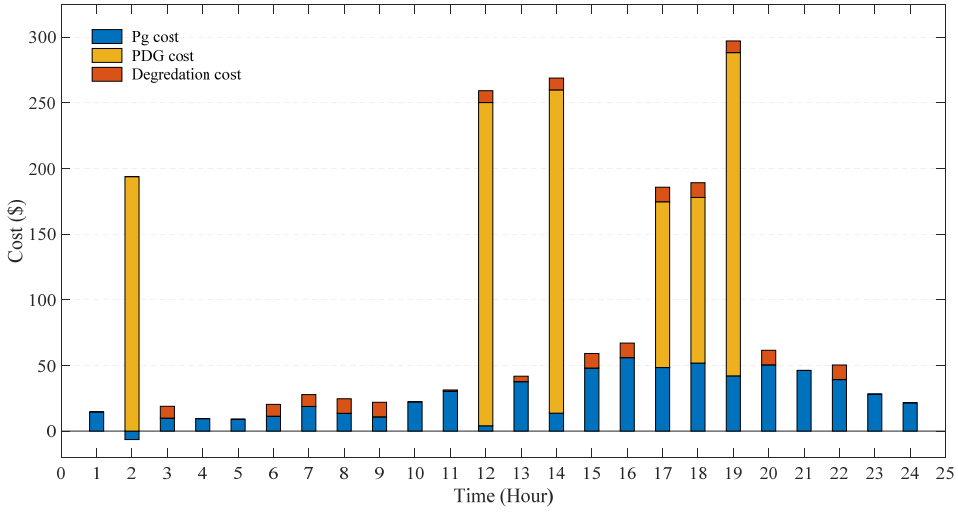


Table 10 The average hourly cost

Energy source	*IPSO*	*DQN*	*Error*
Grid	26.25	27.7	5%
DG	49.368	49.121	0.5%
Degradation	6.166	6.703	8%
Grid+DG+Degredation	81.78	83.5	2%

Figure 14 The entire hourly cost using IPSO (see online version for colours)

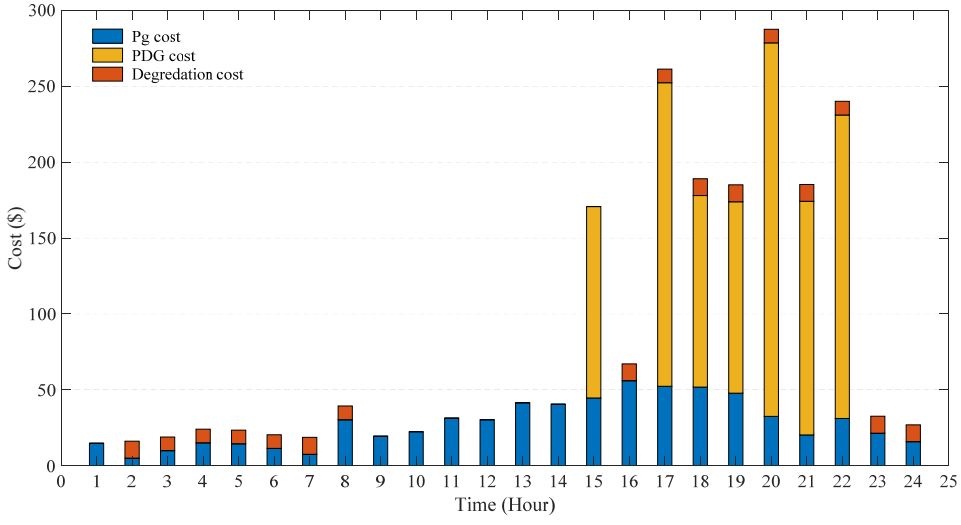


Consequentially, the entire hourly average costs are contiguous and equal to \$81.78/h and \$83.5/h in the two methods with an error ratio equal to 2%. Figures 14 and 15 illustrate the entire daily cost for Pg, PDG, and degradation cost using DQN and IPSO respectively, Table 11 shows the details that lead to the same conclusion of the previous one, where a large convergence to the benchmark values is observed and the overall error rate is equal to 2%.

Table 11 The overall hourly cost (\$)

Energy source	*IPSO*	*DQN*	*Error*
Pg	630	664.67	5%
PDG	1,184.85	1,179	0.5%
Degradation	148	160.888	8%
Pg+PDG+Degredation	1,962.85	2,004.55	2%

Figure 15 The entire hourly cost using DQN (see online version for colours)



6 Conclusions

In this paper, an intelligent EMS had been proposed to schedule and control the fuel consumption and energy resources in a connected-mode MG in rural areas unlike other MGs that are installed in these areas in islanded mode. The system is formulated as an MDP and a model-free RL-DQN strategy had been adopted to tackle the randomness of the system. A simulation had been performed using actual data of a rural area in Ontario City. The fuel supplying to DG is based on the principle of the take-or-pay and EMS was designed to work intelligently to consume the fuel along the time horizon according to this conception. The results obtained using DQN are confirmed and verified by comparing them with benchmark results obtained using the model-based IPSO. Two active days had been elected to apply the optimum policy one from the training

profile and the other from the testing profile. It is observed that the result of the proposed method is near the optimum point of IPSO policy. The energy cost and energy consumption factors had been adopted to measure the efficiency of the proposed method. From the results, it is noted a small error ratio of energy consumption does not exceed 4% using the two methods depending on the energy consumption factor, at the same time, the error ratio of the average hourly cost is 0.125% and for daily cost is 2% in the two methods under energy cost factor consideration. Also, it is noticed that the entire specified fuel is consumed and the limit of PCC is not violated at each time slot, so this reflects the efficiency of the proposed algorithm. It is clear from this paper that the used learning method is more effective in energy management in real-time.

References

Aklo, N.J. and Rashid, M.T. (2021) 'Scheduling of diesel generators operation with restricted pcc in microgrid', *Iraqi Journal for Electrical and Electronic Engineering*, DOI: 10.37917/ijeee.17.2.13.

Al Hadi, A., Silva, C.A.S., Hossain, E. and Challoo, R. (2020) 'Algorithm for demand response to maximize the penetration of renewable energy', *IEEE Access*, 19 March, Vol. 8, pp.55279–55288.

Albadi, M.H. and El-Saadany, E.F. (2007) 'Demand response in electricity markets: an overview', *IEEE Power Engineering Society General Meeting*.

Anglani, N., Oriti, G. and Colombini, M. (2017) 'Optimized energy management system to reduce fuel consumption in remote military microgrids', *IEEE Trans. on Industry Applications*, Vol. 53, No. 6, pp.5777–5785.

Baldwin, C.J., Dale, K.M. and Dittrich, R.F. (1959) 'A study of the economic shutdown or generating units in daily dispatch', *Trans. of the American Institute of Electrical Engineers. Part III: Power Apparatus and Systems*, Vol. 78, No. 4, pp.1272–1282.

Bucolo, M., Buscarino, A., Famoso, C., Fortuna, L. and Frasca, M. (2019) 'Control of imperfect dynamical systems', *Nonlinear Dynamics*, Vol. 98, No. 4, pp.2989–2999.

Bui, V.H., Hussain, A. and Kim, H.M. (2019) 'Q-learning-based operation strategy for community battery energy storage system (CBESS) in microgrid system', *Energies*, Vol. 12, No. 9, pp.1789–1806.

Bui, V-H., Hussain, A. and Kim, H-M. (2020) 'Double deep Q-learning-based distributed operation of battery energy storage system considering uncertainties', *IEEE Trans. on Smart Grid*, Vol. 11, No. 1, pp.457–469.

Busoniu, L., Babuska, R., De Schutter, B. and Ernst, D. (2010) 'Reinforcement learning and dynamic programming using function approximators', CRC Press, Boca Raton, FL, USA.

Clavier, J., Bouffard, F., Rimorov, D. and Joós, G. (2015) 'Generation dispatch techniques for remote communities with flexible demand', *IEEE Trans. on Sustainable Energy*, Vol. 6, No. 3, pp.720–728.

Dillon, T.S., Edwin, K.W., Kochs, H-D. and Taud, R.J. (1978) 'Integer programming approach to the problem of optimal unit commitment with probabilistic reserve determination', *IEEE Trans. on Power Apparatus and Systems*, November/December, Vol. PAS-97, No. 6, pp.2154–2166.

Du, Y. and Li, F. (2020) 'Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning', *IEEE Trans. on Smart Grid*, Vol. 11, No. 2, pp.1–11.

Elkazaz, M., Sumner, M. and Thomas, D. (2020) 'Energy management system for hybrid PV-wind-battery microgrid using convex programming, model predictive and rolling horizon predictive control with experimental validation', *International Journal of Electrical Power & Energy*, February, Vol. 115.

El-Sharkh, M.Y., Rahman, A. and Alam, M.S. (2010) 'Short term scheduling of multiple grid-parallel PEM fuel cells for microgrid applications', *International Journal of Hydrogen Energy*, October, Vol. 35, No. 20, pp.11099–11106.

Fioriti, D., Giglioli, R. and Poli, D. (2017) 'Short-term operation of a hybrid minigrid under load and renewable production uncertainty', *IEEE Global Humanitarian Technology Conference (GHTC)*.

Foruzan, E., Soh, L-K. and Asgarpoor, S. (2018) 'Reinforcement learning approach for optimal distributed energy management in a microgrid', *IEEE Trans. on Power Systems*, September, Vol. 33, No. 5, pp.5749–5758.

Huang, Q., Huang, R., Hao, W., Tan, J., Fan, R. and Huang, Z. (2020) 'Adaptive power system emergency control using deep reinforcement learning', *IEEE Trans. on Smart Grid*, March, Vol. 11, No. 2, pp.1171–1182.

Ismail, M.S., Moghavvemi, M. and Mahlia, T.M.I. (2013) 'Techno-economic analysis of an optimized photovoltaic and diesel generator hybrid power system for remote houses in a tropical climate', *Energy Conversion and Management*, May, Vol. 69, pp.163–173.

Jiang, B. and Fei, Y. (2015) 'Smart home in smart microgrid: a cost-effective energy ecosystem with intelligent hierarchical agents', *IEEE Trans. on Smart Grid*, January, Vol. 6, No. 1, pp.3–13.

Kaabeche, A. and Ibtiouen, R. (2014) Techno-economic optimization of hybrid photovoltaic/wind/diesel/ battery generation in a stand-alone power system', *Solar Energy*, May, Vol. 103, pp.171–182.

KEMA Inc. (2014) *Microgrids – Benefits, Barriers and Suggested Policy Initiatives for the Commonwealth of Massachusetts*, Burlington, MA, USA.

Kim, B-G., Zhang, Y., van der Schaar, M. and Lee, J-W. (2016) 'Dynamic pricing and energy consumption scheduling with reinforcement learning', *IEEE Trans. on Smart Grid*, September, Vol. 7, No. 5, pp.2187–2198.

Kumar, A.B.R., Vemuri, S., Ebrahimzadeh, P. and Farahbakhshian, N. (1984) 'Fuel resource scheduling – the long-term problem', *IEEE Power Engineering Review*, Vol. PER-4, No. 7, pp.145–151.

Lee, F.N., Liao, J. and Breipohl, A.M. (1992) 'Adaptive fuel allocation using pseudo fuel prices', *Trans. on Power Systems*, May, Vol. 7, No. 2, pp.487–496.

Liang, S. and Srikant, R. (2017) 'Why deep neural networks for function approximation?', *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, April, Toulon, France, pp.1–17.

Liu, W., Zhuang, P., Liang, H., Peng, J. and Huang, Z. (2018) 'Distributed economic dispatch in microgrids based on cooperative reinforcement learning', *IEEE Trans. Neur Net Learn Syst.*, June, Vol. 29, No. 6, pp.2192–2203.

Mandal, S., Das, B.K. and Hoque, N. (2018) 'Optimum sizing of a stand-alone hybrid energy system for rural electrification in Bangladesh', *Journal of Cleaner Production*, 1 November, Vol. 200, pp.12–27.

Michaelson, D., Mahmood, H. and Jiang, J. (2018) 'Reduction of forced outages in islanded microgrids by compensating model uncertainties in PV rating and battery capacity', *IEEE Power and Energy Technology Systems Journal*, Vol. 5, No. 4, pp.129–138.

Oliehoek, F.A. (2012) *Decentralized POMDPs in Reinforcement Learning: State-of-the-Art (Adaptation, Learning, and Optimization)*, Ch. 15, Springer-Verlag, Berlin, Germany.

Qiu, X., Nguyen, T.A. and Crow, M.L. (2016) 'Heterogeneous energy storage optimization for microgrids', *IEEE Trans. on Smart Grid*, May, Vol. 7, No. 3, pp.1453–1461.

Sutton, R.S. and Barto, A.G. (2014) *Reinforcement Learning: An Introduction*, 2nd ed., A Bradford Book, The MIT Press, Cambridge, Massachusetts.

Sutton, R.S. and Barto, A.G. (2018) *Reinforcement Learning: An Introduction*, October, MIT Press, Cambridge.

Thirugnanam, K., Kerk, S.K., Yuen, C., Liu, N. and Zhang, M. (2018) 'Energy management for renewable microgrid in reducing diesel generators usage with multiple types of battery', *IEEE Trans. on Industrial Electronics*, Vol. 65, No. 8, pp.6772–6786.

Tong, S.K. and Shahidehpour, S.M. (1990) 'An innovative approach to generation scheduling in large-scale hydro-thermal power systems with fuel constrained units', *IEEE Trans. on Power Systems*, May, Vol. 5, No. 2, pp.665–673.

Tushar, W., Yuen, C., Huang, S., Smith, D.B. and Poor, H.V. (2016) 'Cost minimization of charging stations with photovoltaics: an approach with EV classification', *IEEE Trans. on Intelligent Transportation Systems*, Vol. 17, No. 1, pp.156–169.

Van Meeteren, H.P. (1984) 'Scheduling of generation and allocation of fuel using dynamic and linear programming', *IEEE Power Engineering Review*, Vol. PER-4, No. 7, pp.26–27.

Venayagamoorthy, G.K., Sharma, R.K., Gautam, P.K. and Ahmadi, A. (2016) 'Dynamic energy management system for a smart microgrid', *IEEE Trans. on Neural Networks and Learning Systems*, August, Vol. 27, No. 8, pp.1643–1656.

Wang, Z., Chen, B., Wang, J., Begovic, M.M. and Chen, C. (2015) 'Coordinated energy management of networked microgrids in distribution systems', *IEEE Trans. on Smart Grid*, January, Vol. 6, No. 1, pp.45–53.

Wu, H., Zhuang, H., Zhang, W. and Ding, M. (2016) 'Optimal allocation of microgrid considering economic dispatch based on hybrid weighted bilevel planning method and algorithm improvement', *International Journal of Electrical Power & Energy Systems*, February, Vol. 75, pp.28–37.

Zhang, Q., Dehghanpour, K., Wang, Z. and Huang, Q. (2020) 'A learning-based power management method for networked microgrids under incomplete information', *IEEE Trans. on Smart Grid*, Vol. 11, No. 2, pp.1–12.

Zhang, Q., Lin, M., Yang, L.T., Chen, Z., Khan, S.U. and Lia, P. (2019) 'Double deep Q-learning model for energy-efficient edge scheduling', *IEEE Transactions on Services Computing*, Vol. 12, No. 5, pp.739–749.

Zhu, L., Han, J., Peng, D., Wang, T., Tang, T. and Charpentier, J-F. (2014) 'Fuzzy logic based energy management strategy for a fuel cell/battery/ultracapacitor hybrid ship', *Proc. 1st Int. Conf. Green Energy*, March, pp.107–112.

Nomenclature

C^{DG}	Energy cost of the diesel generator
C^{deg}	Degradation cost of battery
C_t^g	Fuel cost
C_t^g	Energy cost of the grid at time slot t
$E_{bat,\max}$	Maximum capacity of battery
E_t	Energy storage of battery at time slot t
F	The consumed fuel in the generator
$F^{\max}$	Maximum capacity of the fuel tank
P_t^{ch}	Rated power charged to at time slot t
$P_t^{ch,\max}$	Maximum rated power charging at time slot t
P_t^{dis}	Rated power discharged at time slot t
$P_t^{dis,\max}$	Maximum rated power discharging at time slot t
P_t^{DG}	Generated power from DG at time slot t
$P_t^{DG,\min}$	Minimum power allowed from the generator
P_t^g	Imported power from the grid
$P^{DG,rated}$	Rated power of the generator
$P_{pv}^{\max}$	PV power generated at time slot t
$P_{pv}^{\max}$	PV system rating
SOC_t	Energy ratio saved in the battery at time slot t
$SOC^{\max}$	Minimum energy ratio saved in the battery
$SOF^{\max}$	Maximum fuel capacity ratio of the tank
T_c	PV cell temperature
T_t	Ambient temperature forecast at time slot t
ρ^{deg}	Battery degradation factor
η^{ch}	Efficiency of charging energy
η^{dis}	Efficiency of discharging energy