Deep Reinforcement Learning for Energy Microgrids Management Considering Flexible Energy Sources

Nikita Tomin, *Alexey Zhukov*, and Alexander Domyshev

Melentiev Energy Systems Institute SB RAS, Irkutsk, Russia

Abstract. The problem of optimally activating the flexible energy sources (short- and long-term storage capacities) of electricity microgrid is formulated as a sequential decision making problem under uncertainty where, at every time-step, the uncertainty comes from the lack of knowledge about future electricity consumption and weather dependent PV production. This paper proposes to address this problem using deep reinforcement learning. To this purpose, a specific deep learning architecture has been used in order to extract knowledge from past consumption and production time series as well as any available forecasts. The approach is empirically illustrated in the case of off-grid microgrids located in Belgium and Russia.

1 Introduction

An electricity microgrid (MG) is an energy system consisting of local electricity generation, local loads (or energy consumption) and storage capacities. MGs often face difficulties in supplying demand due to the lack of sufficient energy generation sources.

This problem is caused by the uncertain nature of renewable energy sources (such as wind and photovoltaic (PV) generations), market prices as well as loads with new options (such as heat pumps, e-vehicles, storage systems) that lead to difficulties in ensuring power quality and in balancing generation and consumption. To tackle these problems, MGs should be managed by an energy management system (EMS) that facilitates the minimization of operational costs, emissions and peak loads while satisfying the MG technical constraints (Shayeghi et al., 2019).

This paper has introduced a deep reinforcement learning (RL) architecture for addressing the problem of operating an electricity MG in a stochastic environment. We consider MGs that are provided with flexibility options such as different types of storage devices in order to be able to address both short- and long-term fluctuations of electricity production using PV panels (typically, batteries for short-term fluctuations, and hydrogen/fuel cells or/and diesel generator for long-term fluctuations).

Distinguishing short- from long-term storage is mainly a question of cost: batteries are currently too expensive to be used for addressing seasonal variations. Energy MGs face a dual stochastic-deterministic structure: one of the main challenge to meet when operating MGs is to find storage strategies capable of handling uncertainties related to future electricity production and consumption; besides this, MGs also have the characteristics that their dynamics deterministically reacts to storage management actions (Francois-Lavet et al., 2016).

This paper is organized as follows: Section 2 details the energy MG management problem considering flexible options sources. Section 3 describes the deep RL framework. Section 4 introduces our Deep RL structure dedicated to MG management, as well as empirical results corresponding to the case of off-grid MGs located in Belgium and Russia. Section 5 concludes.

2 Energy Microgrids Management

2.1 Microgrid concept

Distributed generations eliminate the need for the transmission system by being installed near the customers (Aboli et al., 2019). Integration and control of distributed generations s along with storage devices and flexible loads can constitute a low voltage distribution network, called a MG, which can be operated in isolated or grid-connected mode (Sedighizadeh et al., 2019). The generic concept of a MG is shown in Figure 1.

In the grid-connected mode, ancillary services can be provided by trading activity between the MG and the main grid (Table 1, 2). Other possible revenue streams exist (Karavas et al., 2015). In the islanded mode, the real and reactive power generated within the MG, including that provided by the energy storage system, should be in balance with the demand of local loads. MGs offer an option to balancing the need to reduce carbon emissions while continuing to provide reliable electric energy in

* This work was supported by the Russian Scientific Foundation (No. 19-49-04108) under the project ”Development of Innovative Technologies and Tools for Flexibility Assessment and Enhancement of Future Power Systems”.

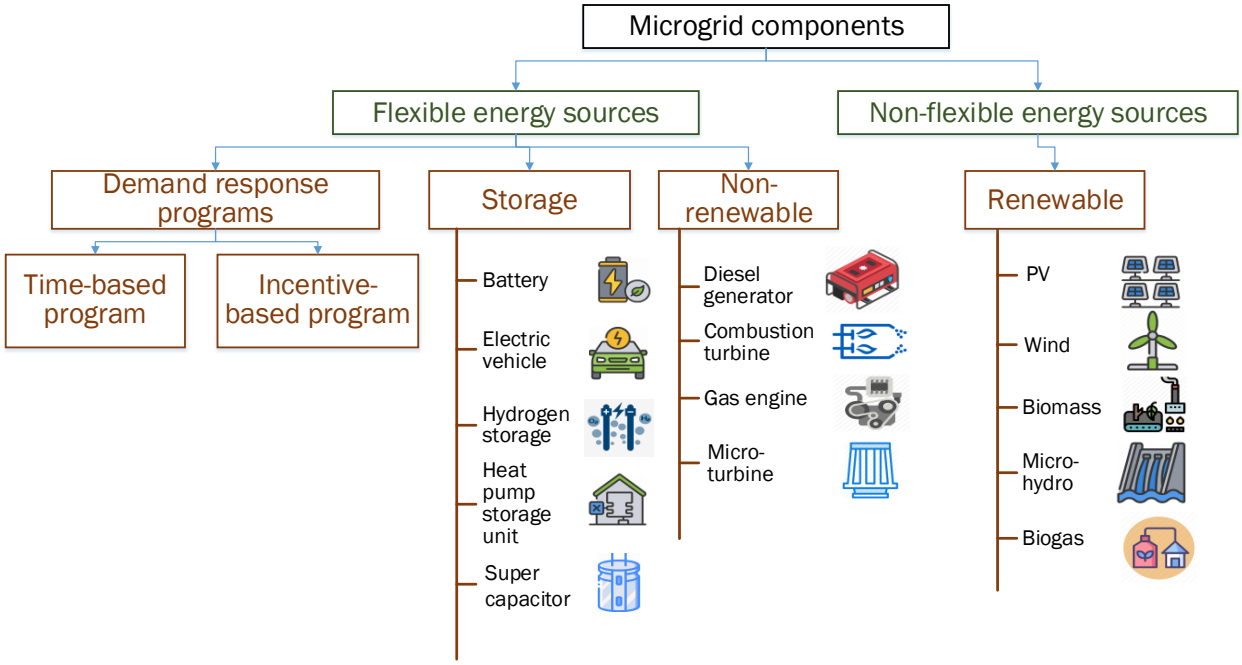


Fig. 1. Microgrid components.

periods of time that renewable sources of power are not available.

Table 1. The creation mechanisms of a grid-tied MG

Function	Description	Battery storage system
Energy markets	Decide on the price your are willing to pay/sell	++
Ancillary services	Sell services to the grid	++
Peak reduction	Through local and community optimization	++
UPS functionality	Operate in islanded mode	++
Efficiency	Through optimized load and generation management	

Table 2. Advantages for the public grid

Function	Description	Battery storage system
Peak reduction /flow management	Momentarily set constraints to the MG	++
Voltage support	Reactive power flexibility of battery storage and PV	++
Phase balancing	Using storage DC buffer	++
Power factor correction	Flexibility of inverters	++
Frequency support	Primary or secondary reserve	++

A MG may transition between these two modes because of scheduled maintenance, degraded power modifying energy flow through MG components, MGs facilitate the integration of renewable energy generation such as PV, wind and fuel cell generations without requiring re-design of the national distribution system (Sfikas et al., 2015; Logenthiran et al., 2011). Modern optimization methods can also be incorporated into the MG EMS to improve efficiency, economics, and resiliency.

2.2 Intelligent EMS design

The aim of an EMS is to determine the optimal use of distributed generations in order to feed the electrical loads (Li et al., 2019). An EMS can be operated in two modes, namely centralized and decentralize. In the centralized mode, the central controller aims to optimize the microgrid power exchanged based on the market prices and security constraints. In the decentralized mode, MGs and controllable loads have more degree of freedom (Dou et al., 2015). As a result, the MG components are considered to be intelligent and try to maximize the revenue of the microgrid by communicating with each other (Katiraei et al., 2008). The initial duty of EMS in both centralized and decentralized mode is to ensure MG of providing load-generation balance (Theo et al., 2017).

A standard EMS can usually provide minimal management function, first of all, energy monitoring and fixed rules for storage operation. An intelligent MG EMS can exploit data to make the MG flexible, robust, and extract the maximum of value and has a community management feature (Fig 2). Functional modules that exploit data from MG is provided in Fig. 3. Let consider some blocks in more details.

Energy management in MGs is typically formulated as an offline optimization problem for day-ahead scheduling

by previous studies (Francois-Lavet et al., 2016). Operational planning for MG can cover the following functions:

- Optimize operation by anticipating on the evolution of load, generation and prices, taking into account the technical constraints of MG
- Important to plan the operation of storage systems, and other devices having a highly "time-coupled" behaviour such as flexible loads, or steerable generators
- Islanded mode: take preventive decisions to maintain the power to critical loads as long as possible.

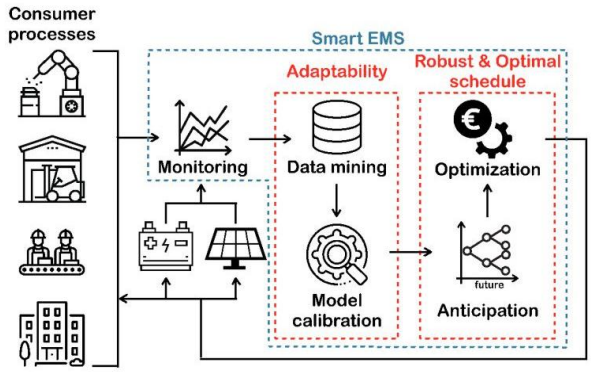


Fig. 2. Intelligent MG EMS design

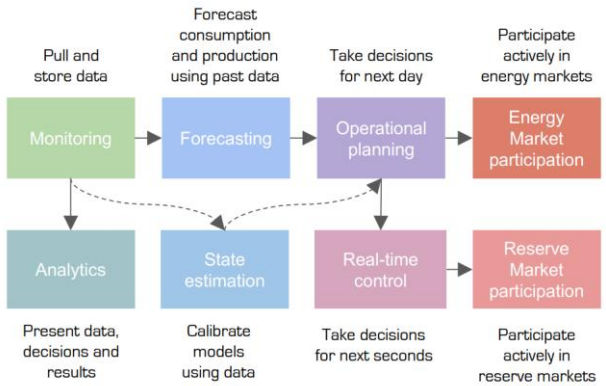


Fig. 3. Functional modules for energy microgrids management

Recently, there have been proposes to develop online algorithms for real-time energy management in MGs to optimize the long-term cost, which take into account the uncertainties of the renewables, the demands, and the market. The following real-time actions can be provided:

- Grid-tied mode: implements operational planning decisions, corrects the error and dispatches among flexibility sources and manages storage systems to limit their degradation
- Islanded mode: monitors and dispatches flexibility sources to maintain system frequency and dispatch of hybrid energy storage systems.

Advanced energy/ancillary services market participation for MG can include:

- Optimal bidding in day-ahead market using anticipated load, generation, and prices.
- Adjusting energy exchanges in intra-day market to match changes in load, generation, and prices.
- Exploiting balancing opportunities by reacting to TSO's (Transmission System Operators) signals.
- Providing remunerated flexibility margins that the TSO can activate for balancing purposes.

3 Deep RL for energy microgrids management

Reinforcement learning (RL) gives a machine the ability to learn to take actions (Sutton et al., 2018). The machine takes actions in an environment to optimize a reward signal. In the context of a microgrid that reward signal could be energy cost, the peak of load, or safety - whatever behaviour we want to incentivize (Fig. 4). In a RL context, an agent learns to act using a Markov decision process (MDP) formalism. However, the state space is very large in modern power grids and therefore a typical RL algorithm has no effective to solve. To solve such problem we can use a deep neural network (NN) to model the desired policies, value functions, which therefore is called Deep RL (François-Lavet et al., 2018).

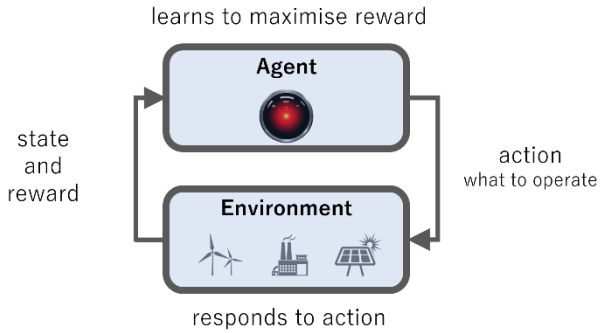


Fig. 4. Agent-environment interaction in RL

3.1 Deep RL solutions for sequential decision making

Optimally operating a MG can be formalized as a partially observable MDP, where the MG is considered as an agent that interacts with its environment. In order to approach the Markov property, the state of the system $s_t \in S$ is made up of an history of features of observations $O_t^i, i \in \{1, \dots, N_f\}$, where $N_f \in N$ is the total number of features. Each O_t^i is represented by a sequence of punctual observations over a chosen history of length h_i: $O_t^i = [o_{t-h^i+1}^i, \dots, o_t^i]$ (the history length may depend on the feature). At each time step, the agent observes a state variable st , takes an action $a_t \in A$ and moves into a state $s_{t+1} \sim P(\cdot|s_t, a_t)$. A reward signal $r_t = \rho(s_t, a_t, s_{t+1})$ is associated to the transition (s_t, a_t, s_{t+1}), where $\rho: S \times A \times S \to R$ is the reward function. We then define the γ-discounted optimal Q-value function:

$$Q^*(s,a) = \max_\pi E\left[\sum_{k=t}^{\infty} \gamma^{k-T} r_k | s_t = s, a_t = a, \pi\right] \quad (1)$$

By analogy (Francois-Lavet et al., 2016), we propose to approximate $Q^*(s,a)$ using a deep NN. We denote by $Q(\cdot,\cdot;\Theta_k)$ the so-called Q-network (Fig.5). Deep NNs offer generalization properties that are adapted to highdimensional sensory inputs such as temporal series. The NN parameters Θ_k may be updated using stochastic gradient descent by sampling batches of transitions (s,a,r,s') in a replay memory, updating the current value $Q(s,a;\Theta_k)$ towards a target value $Y_k^Q = r + \gamma\, arg\max_{a' \in A} Q(s',a',\Theta_k^-)$ where Θ_k^- refers to parameters from some previous Q-network called the target Q-network or deep Q-network (DQN) as introduced in (Mnih et al., 2015). When using the squared-loss, a Q-learning update is obtained as follows:

$$\Theta_{k+1} = \Theta_k + \alpha(Y_k^Q - Q(s,a;\Theta_k))\nabla_{\Theta_k} Q(s,a;\Theta_k) \quad (2)$$

where α is a scalar step size called the learning rate.

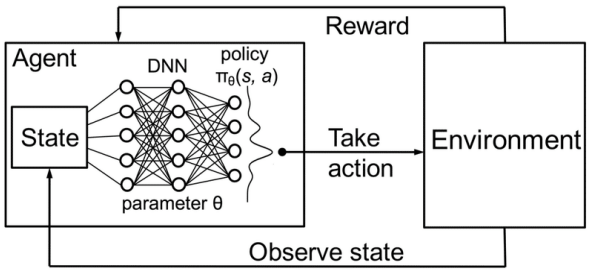


Fig. 5. Deep RL design

3.2 Microgrid: benchmark description

First, note that the MG model described hereafter is fully described in (Gemine et al., 2016). We denote the storage operation state of the microgrid by $s_t^{MG} \in S^{MG}$: it describes the amount of energy in the storage devices. The amount of energy in the battery is denoted by $s_t^B\ [Wh] \in S^B[Wh]$, the amount of energy in the hydrogen tank is denoted by $s_t^{H2}\ [Wh] \in S^{H2}[Wh]$ and the amount of energy density in the diesel generator is denoted by $s_t^{DG}\ [Wh] \in S^{DG}[Wh/kg]$. We introduce $x_B\ [Wh]$ (resp. $x_{H2}\ [Wp]$) as the battery (resp. hydrogen) storage sizing and the diesel generator power $x_{DG}\ [W]$. The variable η_B (resp. ζ_B) denotes the battery discharge (resp. charge) efficiency. Similarly, the electrolysis/fuel cells efficiencies are respectively denoted by η_{H2} (when storing energy) and ζ_{H2} (when delivering energy). The variable ζ_{DG} denotes the diesel generator efficiency.

At every time step, an action $a_t = [a_t^{H2}, a_t^{DG}, a_t^B] \in A_t$, is applied on the system, where a_t^{H2} is the amount of energy transferred into (if positive) or out of (if negative) the hydrogen storage device, similarly a_t^B is the amount of energy transferred into or out of the battery and a_t^{DG} is the amount of energy out of (all negative) the diesel generator. The battery dynamics is given by: $s_{t+1}^B = s_t^B + \eta_t^B a_t^B$ if $a_t^B \geq 0$ and $s_{t+1}^B = s_t^B - a_t^B/\zeta_t^B$ otherwise. Similarly, the hydrogen dynamics is given by: $s_{t+1}^{H2} = s_t^{H2} + \eta_t^{H2} a_t^{H2}$ if $a_t^{H2} \geq 0$ and $s_{t+1}^{H2} = s_t^{H2} - a_t^{H2}/\zeta_t^{H2}$ otherwise. The diesel generator dynamics is given by: $s_{t+1}^{DG} = s_t^{DG} - a_t^{DG}/\zeta_t^{DG}$ for all cases.

The reward function of the system corresponds to the instantaneous operational revenues r_t at time $t \in T$. We used three quantities that are prerequisites to the definition of the reward function: (1) $\phi_t\ [Wh] \in \mathbb{R}^+$ is the electricity generated locally by the PV installation, (2) $d_t\ [Wh] \in \mathbb{R}$ denotes the net electricity demand, which is the difference between the local consumption c_t and the local production of electricity φ_t, (3) $\delta_t [Wh] \in \mathbb{R}$ [represents the power balance within the microgrid, taking into account the contributions of the net electricity demand and the charge or discharge of the storage devices: $\delta_t = -a_t^B - a_t^{H2} - a_t^{DG} - d_t$ (Fig. 6).

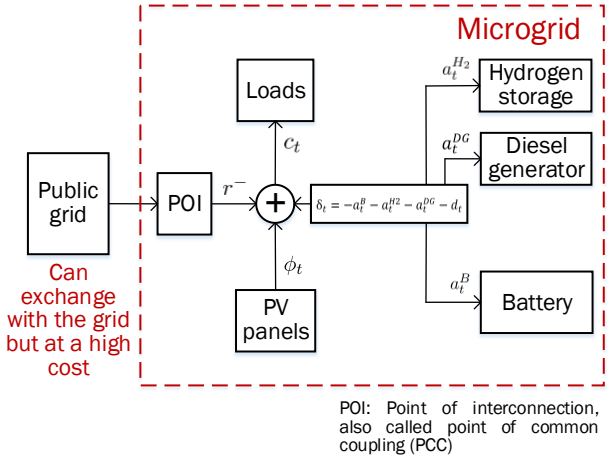


Fig. 6. Schema of the MG featuring PV panels associated with a battery, a hydrogen storage and diesel unit device.

The instantaneous reward signal r_t is obtained by adding the revenues generated by the hydrogen production r^{H2} with the penalties r^- due to the value of loss load: $r_t = r(a_t, d_t) = r^{H2} + r^-(a_t, d_t)$. The penalty r^- is proportional to the total amount of energy that was not supplied to meet the demand: $r^-(a_t, d_t) = k\delta_t$ when $\delta_t < 0$ and null otherwise (k is the cost endured per Wh not supplied within the microgrid), while r_{H2} is given by: $r^{H2}(a_t, d_t) = k^{H2} a_t^{H2}$ (k^{H2} is the revenue/cost per Wh of hydrogen produced/used). In accordance with the problem statement, there is no way to supply energy outside (for public grid) and the system does not receive a reward for this. However, a hybrid PV-hydrogen-diesel system can provide much more substantial maximization of operating revenie.

From the series of rewards r_t, we obtain the operational revenues over year y defined as follows: $M_y = \sum_{t \in \tau_y} r_t$ where τ_y is the set of time steps belonging to year y. Optimizing the operation of the MG requires to determine a sequential decision making strategy that leads to the maximization of M_y (Francois-Lavet et al., 2016).

4 Case studies

The proposed approach is empirically illustrated in the cases of intelligent energy management of off-grid MGs

located in Belgium and Russia. Our examples simulate the operation of a realistic MG that is not connected to the main utility grid (off-grid) and that is provided with PV panels, batteries and hydrogen storage and/or diesel generator. We used and modified the initial source Python code proposed in (Francois-Lavet et al., 2016).

We proposed a DQN architecture where the inputs are provided by the state vector, and where each separate output represents the Q-values for each discretized action. The DQN processes time series thanks to a set of convolutions with 16 filters of 2×1 with stride 1 followed by a convolution with 16 filters of 2×2 with stride 1. The output of the convolutions as well as the other inputs are then followed by two fully connected layers with 50 and 20 neurons and the output layer. The activation function used is the Rectified Linear Unit (ReLU) except for the output layer where no activation function is used.

By starting with a random DQN, we perform at each time step the update given in Eq. 2 and, in the meantime, we fill up a replay memory with all observations, actions and rewards using an agent that follows an ϵ-greedy policy s.t. the policy $\pi(s) = \max_{a \in A} Q(s, a; \Theta_k)$ is selected with a probability $1 - \epsilon$, and a random action (with uniform probability over actions) is selected with probability ϵ. We use a decreasing value of ϵ over time. During the validation and test phases, the policy $\pi(s) = \max_{a \in A} Q(s, a; \Theta_k)$ is applied (with $\epsilon = 0$).

The state of the DQN agent is made up of a history of two to four punctual observations:

- Harging state of the short term storage (0 is empty, 1 is full)
- Production φ_t and consumption c_t
- Distance to equinox
- Predictions of future production for the next 24 hours and 48 hours

4.1 Case 1: PV hydrogen hybrid system

The DQN-agent can either choose to store in the long term storage (hydrogen storage) or take energy out of it while the short term storage (battery) handle at best the lack or surplus of energy by discharging itself or charging itself respectively (Fig. 7). We consider three discretized actions: (1) discharge at full rate the hydrogen storage, (2) keep it idle or (3) charge it at full rate. Whenever the short term storage is empty and cannot handle the net demand a penalty (negative reward) is obtained equal to the value of loss load set to 2euro/kWh.

We consider the case of a residential electricity consumer (average of 18kWh/day) located in Belgium operating an off-grid MG (Fig. 8). The size of the battery is $x_B = 15\ kWh$, the instantaneous power of the hydrogen storage is $x_{H2} = 1.1\ kW$ and the peak PV power generation is $x_{PV} = 12\ kWp$. The cost k endured per kWh not supplied within the MG is set to 2 euro/kWh. Other MG parameters are taken from (Gemine et al., 2016).

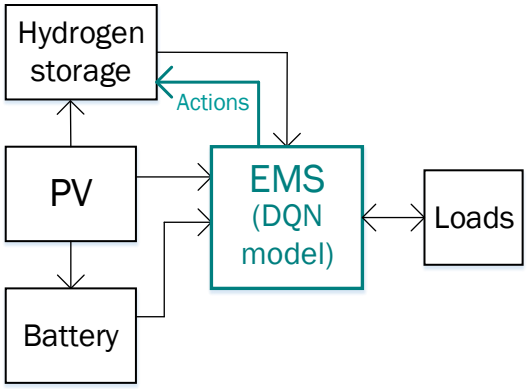


Fig. 7. Schema of the MG featuring PV panels associated with a battery, a hydrogen storage and diesel unit device.

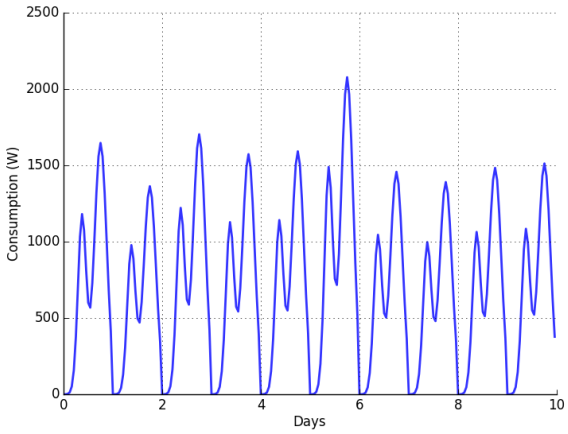


Fig. 8. Representative residential consumption profile for Belgium off-grid MG

Solar irradiance varies throughout the year depending on the seasons, and it also varies throughout the day depending on the weather and the position of the sun in the sky relative to the PV panels. The main distinction between these profiles is the difference between summer and winter PV production. In particular, production varies with a factor 1:5 between winter and summer as can be seen from the measurements of PV panels production for a residential customer located in Belgium in the Figures 9 and 10.

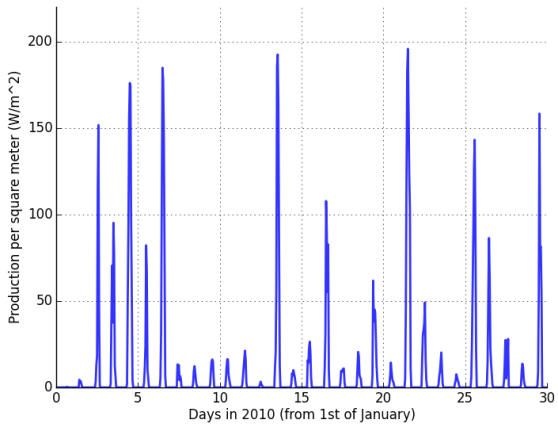


Fig. 9. Typical PV production in winter

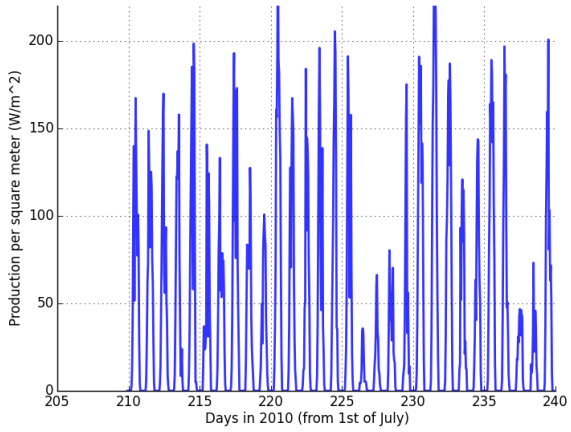
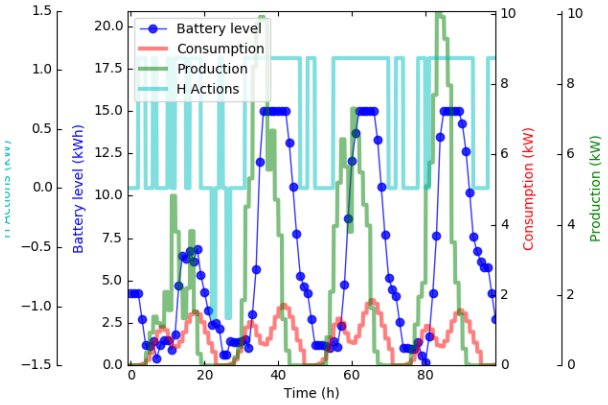
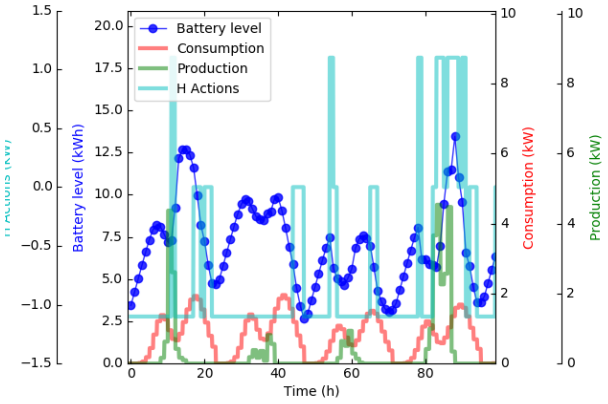


Fig. 10. Typical PV production in summer

The typical behaviour of the policy, π^* is illustrated in Figure 11 (test data). Since the microgrid has no information about the future, it builds up (during the night) a sufficient reserve in the short-term storage device so as to be able to face the next day consumption without suffering too much loss load. It also avoids wasting energy (when the short term storage is full) by storing in the long-term storage device whenever possible.



(a) Typical policy during summer



(b) Typical policy during winter

Fig. 11. Computed policy with minimal information available to the agent. H action = 0 means discharging the hydrogen reserve at maximum rate; H action = 1 means doing nothing with the hydrogen reserve; H action = 2 means building up the hydrogen reserve at maximum rate.

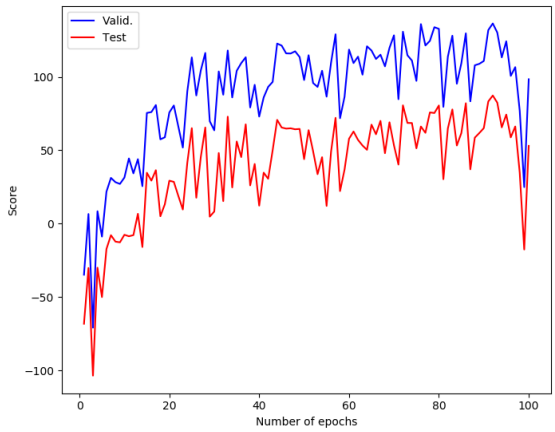


Fig. 12. Operational revenue M_y for the Case 1

We report in Figure 12 the operational revenue on the test and validation data M_y for the Case 1 as a function of DQN training process. Best DQN obtained after 92 epochs, with validation score 136.46 euro/year. Test score of this neural network is 87.34 euro/year.

4.2 Case 2: PV diesel hybrid system

The advantage of such a system is that it needs low investment cost. However, the main disadvantages are that it needs to supply fuel for the operation of the generator and the surplus energy during the good season is not stored (Fig. 13).

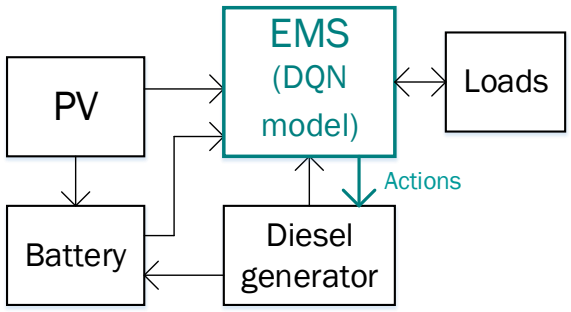


Fig. 13. Schema of the MG featuring PV panels associated with a battery, a hydrogen storage and diesel unit device.

We consider the Case 2 of a residential electricity consumer (average of 48kWh/day) located in Yakutia, Russia operating an off-grid MG. Retrospective data (time interval from 2005 to 2019) on the total solar radiation were used as baseline information. Solar radiation was recorded at a meteorological station located in this settlement. Electrical load shown in Fig. 14 was built according to the real data of typical days relative to each month. The size of the battery is $x_B = 384\ kWh$, the power of the diesel generator is $x_{DG} = 100\ kW$ and the peak PV power generation is $x_{PV} = 75\ kWp$. The cost k endured per kWh not supplied within the MG is set to 2 euro/kWh. The MG parameters are taken from (Sidorov et al., 2019).

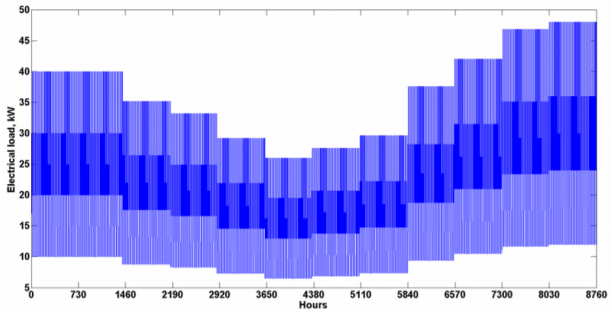
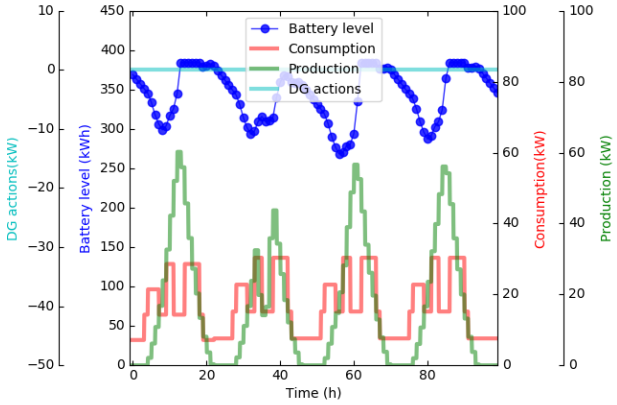


Fig. 14. Representative residential consumption profile for Russia off-grid MG

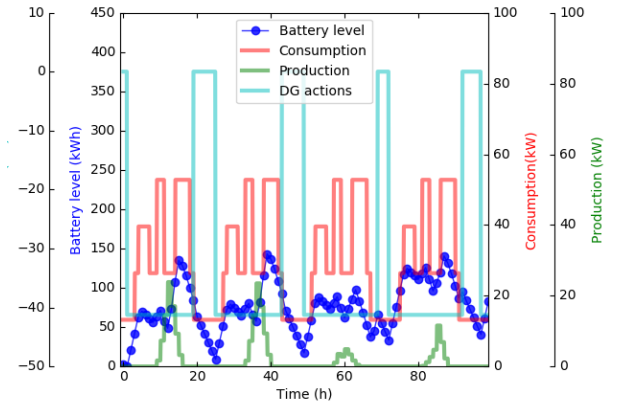
The state of the DQN agent is made up of a history of two to four punctual observations, which the same as in the Case 1. Three actions are possible for the DQN-agent:

- DG Action -2 corresponds to turning ON at full capacity (Emergency Standby Power Mode)
- DG Action -1 corresponds to turning ON at 75% capacity (Prime Power Mode)
- DG Action 0 corresponds to keeping it idle.

The typical behaviour of the policy is illustrated in Figure 15 (test data).



(a) Typical policy during summer



(b) Typical policy during winter

Fig. 15. Computed policy with extended information available to the agent. DG action = 0 means keeping it idle; DG action = -40 means turning ON at 75% capacity (Prime Power Mode); DG action = -60 means turning ON at full capacity (Emergency Standby Power Mode).

It can be seen from Figure 15 that during summer, the diesel generator is idle, and the hydrogen storage is charging; in winter - both devices work to provide a power balance for the MG. Figure 16 shows the operational revenue on the test and validation data M_y for the Case 2 as a function of DQN training process. Best DQN obtained after 5 epochs, with validation score -100.056 euro/year. Test score of this neural network is -100.44 euro/year.

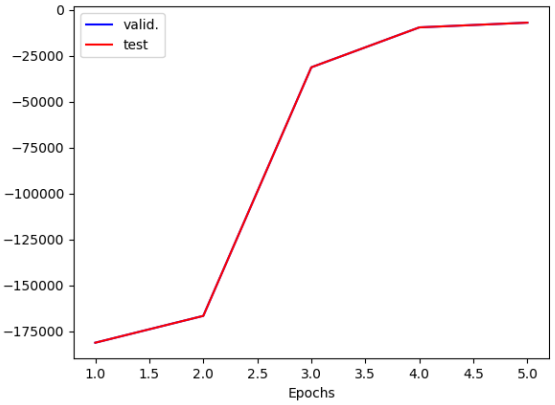


Fig. 16. Operational revenue M_y for the Case 2

Negative operational revenue M_y for the optimal policy π^* here means that analysed PV diesel MG does not have the ability to accumulate excess PV production in a long-term storage. In this case, the main goal is not maximizing operational revenue, but actually minimizing costs.

4.3 Case 3: PV hydrogen diesel hybrid system

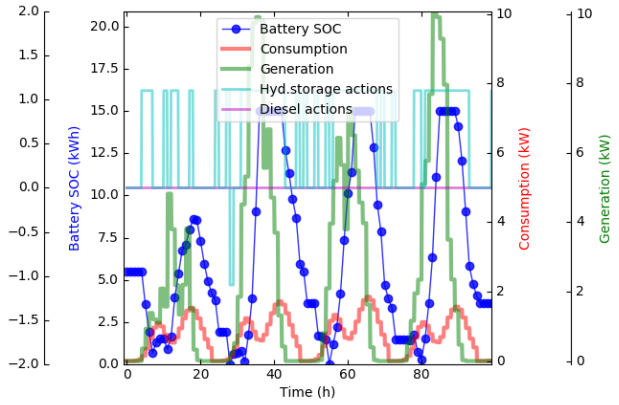
We considered more complicated isolated hybrid system, when DQN-agent can manage two flexibility options - hydrogen storage and diesel generator. We used the data for the Case 1 (off-grid Belgium MG). The design of such hybrid system poses a more complex problem of optimization (Dufo-Lopez et al., 2008).

The state of the DQN agent is made up of a history of two to four punctual observations, which the same as in the Cases 1 and 2.

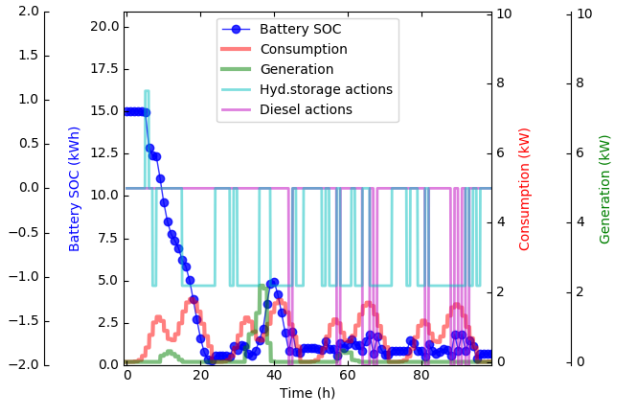
Six actions are possible for the DQN-agent:

- DG action/H action = 0 means keeping it idle or to do nothing;
- DG action = -2 means turning ON at full capacity;
- DG action = -1 means turning ON at 75% capacity;
- H action = -1 means discharging the hydrogen storage;
- H action = 1 means charging the hydrogen storage

The typical behaviour of the policy, π^* for the Case 3 is illustrated in Figure 17 (test data). It can be seen that during summer the diesel generator is idle, as well as the hydrogen storage is charging. However during winter, both long-term storages work to help PV panels in providing demand. At the same time, the hydrogen storage is switched on (discharged) more often than the more expensive diesel generator.



(a) Typical policy during summer



(b) Typical policy during winter

Fig. 17. Computed policy with extended information available to the agent. DG action/H action = 0 means keeping it idle or to do nothing; DG action = -2 means turning ON at full capacity; DG action = -1 means turning ON at 75% capacity; H action = -1 means discharging the hydrogen storage; H action = 1 means charging the hydrogen storage

We report in Figure 18 the operational revenue on the test and validation data M_y for the Case 3 as a function of DQN training process. Best DQN obtained after 91 epochs, with validation score 79.59 euro/year. Test score of this neural network is 26.44 euro/year.

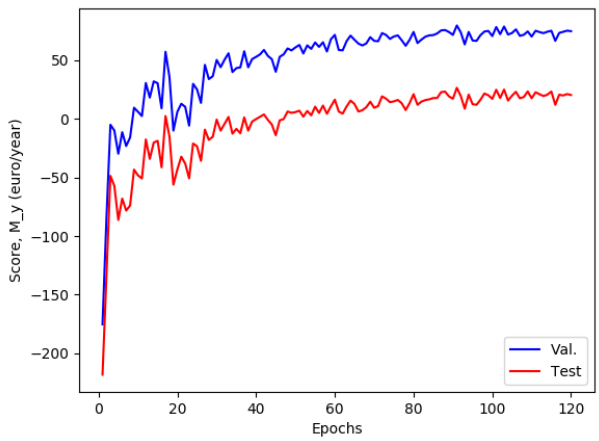


Fig. 18. Operational revenue M_y for the Case 3

It is interesting to note that for the Case 3, operating revenues M_y are lower than for the Case 1. This is the expected result, since the addition of a diesel generator leads to additional fuel costs.

5 Conclusions

Relying on traditional mathematical models to manage the distributed flexibility in real time for MG is not effective, since the computing power to process them would be extremely large. The probable solution is to use the combination of two modelling approaches (optimization problem of flexibility options + realistic stochastic inputs) can be combined with an extensive simulation based on deep reinforcement learning.

This paper has introduced a deep reinforcement learning architecture for addressing the problem of operating an electricity MG in a stochastic environment. Experimental results illustrate the fact that the NN representation of the value function efficiently generalizes the policy to situations corresponding to unseen configurations of electricity demand and solar irradiance.

References

1. H. Shayeghi, E. Shahryari, M. Moradzadeh, P. Siano, A Survey on Microgrid Energy Management Considering Flexible Energy Sources, Energies, **12**, 2156 (2019)
2. V. Francois-Lavet et al., Deep Reinforcement Learning Solutions for Energy Microgrids Management, in European Workshop on Reinforcement Learning, (2016)
3. R. Aboli, M. Ramezani, H. Falaghi, Joint optimization of day-ahead and uncertain near real-time operation of microgrids, Int. J. Electr. Power Energy Syst., **107**, 34–46, (2019)
4. M. Sedighizadeh, M. Esmaili, A. Jamshidi, M.-H. Ghaderi, Stochastic multi-objective economic-environmental energy and reserve scheduling of microgrids considering battery energy storage system, Int. J. Electr. Power Energy Syst., 2019, **106**, 1–16, (2019)
5. C.-S. Karavas, G. Kyriakarakos, K.G. Arvanitis, G. Papadakis, A multi-agent decentralized energy management system based on distributed intelligence for the design and control of autonomous polygeneration microgrids, Energy Convers. Manag. **103**, 166–179, (2015)
6. E.E. Sfikas, Y.A. Katsigiannis, P.S. Georgilakis, Simultaneous capacity optimization of distributed generation and storage in medium voltage microgrids. Int. J. Electr. Power Energy Syst., 2015, **67**, 101–113, (2015)
7. T. Logenthiran, D. Srinivasan, A.M. Khambadkone, Multi-agent system for energy resource scheduling of integrated microgrids in a distributed system, Electr. Power Syst. Res., **81**, 138–148, (2011)

8. B. Li, R, Roche, D. Paire, A. Miraoui, A price decision approach for multiple multi-energy-supply microgrids considering demand response, Energy, **167**, 117–135, (2019)
9. C. Dou, et al., Decentralised coordinated control of microgrid based on multi-agent system. IET Gener. Transm. Distrib., **9**, 2474–2484, (2015)
10. F. Katiraei, R. Iravani, N. Hatziargyriou, A. Dimeas, A. Microgrids management. IEEE Power Energy Mag., **6**, 54–65, (2008)
11. W.L. Theo, et al., Review of distributed generation (DG) system planning and optimisation techniques: Comparison of numerical and mathematical modelling methods. Renew. Sustain. Energy Rev., **67**, 531–573, (2017)
12. R.S. Sutton, A.G. Barto, *Introduction to Reinforcement Learning* (MA: MIT Press, Cambridge, 2018).
13. V. Francois-Lavet, P. Henderson, R. Islam, M.G. Bellemare, J. Pineau, An Introduction to Deep Reinforcement Learning, Foundations and Trends in Machine Learning, **11**(3-4), (2018)
14. V. Mnih, K. Kavukcuoglu, D. Silver, et al. Human-level control through deep reinforcement learning, Nature, **518**(7540), 529–533, (2015)
15. Q. Gemine, V. François-Lavet, D. Ernst, R. Fonteneau. Towards the minimization of the levelized energy costs of microgrids using both long-term and short-term storage devices, Smart Grid: Networking, Data Management, and Business Models, 295–319, (2016).
16. D. Sidorov, I. Muftahov, N. Tomin et al., A Dynamic Analysis of Energy Storage with Renewable and Diesel Generation Using Volterra Equations, IEEE Trans. on Industrial Informatics, **14**(8), (2019)
17. R. Dufo-Lopez, J.L. Bernal-Agustin, Multi-objective design of PV–wind–diesel–hydrogen–battery systems, Renewable Energy, **33**(12), 2559-2572, (2008)