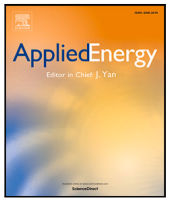
Applied Energy 368 (2024) 123435

Contents lists available at ScienceDirect

Applied Energy

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



Energy management of a microgrid considering nonlinear losses in batteries through Deep Reinforcement Learning

David Domínguez-Barbero *, Javier García-González, Miguel Á. Sanz-Bobi, Aurelio García-Cerrada

Institute for Research in Technology, ICAI School of Engineering, Comillas Pontifical University, Santa Cruz de Marcenado, 26, 28015, Madrid, Spain

ARTICLE INFO

Keywords:
Deep Reinforcement Learning
Energy management system
Energy savings
Isolated microgrid
Nonlinear battery model

ABSTRACT

The massive deployment of microgrids could play a significant role in achieving decarbonization of the electric sector amid the ongoing energy transition. The effective operation of these microgrids requires an Energy Management System (EMS), which establishes control set-points for all dispatchable components. EMSs can be formulated as classical optimization problems or as Partially-Observable Markov Decision Processes (POMDPs). Recently, Deep Reinforcement Learning (DRL) algorithms have been employed to solve the latter, gaining popularity in recent years. Since DRL methods promise to deal effectively with nonlinear dynamics, this paper examines the Twin-Delayed Deep Deterministic Policy Gradient (TD3) performance – a state-of-the-art method in DRL – for the EMS of a microgrid that includes nonlinear battery losses. Furthermore, the classical EMS-microgrid interaction is improved by refining the behavior of the underlying control system to obtain reliable results. The performance of this novel approach has been tested on two distinct microgrids – a residential one and a larger-scale grid – with a satisfactory outcome beyond reducing operational costs. Findings demonstrate the intrinsic potential of DRL-based algorithms for enhancing energy management and driving more efficient power systems.

1. Introduction

Transitioning to cleaner energy sources worldwide in order to reduce CO_2 emissions requires increasing the penetration of Renewable Energy Sources (RESs) into the power generation mix. In fact, during the last two decades, the share of renewable generation in the energy mix has increased significantly. For example, in 2004, the share of renewable generation in Europe was 9.6% of the total generation, while it was 21.3% in 2021, and it is targeted to reach at least 42.5% in 2030 [1]. The majority of this renewable energy production comes from Photovoltaic Panels (PVs) and Wind Turbines (WTs), both of which may be installed within the distribution network as Distributed Generators (DGs), close to loads. Additionally, more conventional power generation units are still required during periods of high demand or low renewable generation.

Energy Storage Systems (ESSs) such as batteries or flywheels play a crucial role when RESs are present. In electric power networks, ESSs may increase frequency stability, provide demand flexibility, and store the energy surplus of renewable generation.

A microgrid is a localized, small-scale power grid that can function either independently or in connection with conventional grids. It usually comprises various DGs and ESSs, which are connected in close proximity to the loads they serve. Moreover, they may support well the integration of RESs, thereby contributing to the reduction of dependence on fossil fuels. Fig. 1 illustrates a typical microgrid setup. Microgrids must manage power generation, distribution, and consumption within their defined boundaries. With effective operation, they can enhance reliability, resilience, and energy efficiency, [2,3]. For instance, microgrids can work disconnected from the main grid during outages or emergencies, ensuring a continuous power supply to critical systems, [4]. Furthermore, they are particularly beneficial in remote areas where extending the main grid is not possible.

A microgrid necessitates a sophisticated digital system that controls each component dynamically and continuously. Microgrid control hierarchy typically consists of three levels, each addressing different aspects of the microgrid's operation, [5]. Primary control is responsible for the fast, local regulation of voltage and frequency, ensuring stable and

* Corresponding author.
E-mail addresses: ddominguez@comillas.edu (D. Domínguez-Barbero), javiergg@comillas.edu (J. García-González), masanz@comillas.edu (M.Á. Sanz-Bobi), aurelio@comillas.edu (A. García-Cerrada).

https://doi.org/10.1016/j.apenergy.2024.123435
Received 18 December 2023; Received in revised form 26 April 2024; Accepted 8 May 2024
Available online 18 May 2024

Notation

Sets and indexes

$t \in T$	Time periods from 1 to $\|T\|$ (= 8760 for hours in one year)
$res \in Res$	Renewable Energy Sources installed in the microgrid
$g \in G$	Non-renewable generator installed in the microgrid

Parameters:

Time series:

D_t	Total load of the microgrid at time period t, [kW]
P_t^{pv}	Maximum available PV power generation at time period t, [kW]
P_t^{wt}	Maximum available WT power generation at time period t, [kW]

Model boundaries:

P_{max}^{pv}	Nominal rate of the solar panel, [kW]
P_{max}^{wt}	Nominal rate of the wind turbine, [kW]
P_{max}^{g}	Maximum output power of the generator $g \in G$, [kW]
P_{min}^{g}	Minimum output power of the generator $g \in G$, [kW]
$P_{max}^{b\leftarrow}$	Maximum power the ESS $b \in B$ can charge, [kW]
$P_{max}^{b\rightarrow}$	Maximum power the ESS $b \in B$ can discharge, [kW]
S_{max}^{b}	Maximum energy capacity in the ESS $b \in B$, [kWh]
S_{min}^{b}	Minimum energy capacity in the ESS $b \in B$, [kWh]

Model dynamics:

δ_2^g	Quadratic term of the generator $g \in G$ cost function, [€/kW^2]
δ_1^g	Linear term of the generator $g \in G$ cost function, [€/kW]
δ_0^g	No-load term of the generator $g \in G$ cost function when committed, [€]
η^b	Charge linear efficiency of the ESS $b \in B$ [Unitless]
ζ^b	Discharge linear efficiency of the ESS $b \in B$ [Unitless]
C^{nse}	Cost of Not Supplied Energy, [€/kWh]
Δ_t	Temporal space between two contiguous time periods: t and $t + \Delta t$, [h]

Starting point:

S_0^b	Initial energy stored inside each ESS $b \in B$, [kWh]

Relationships:

$SOC_t^b = S_t^b / S_{max}^b$	State of charge of the ESS at time period t

Variables

EMS set-points:

$\hat{P}_t^g$	EMS set-point given to the non-renewable generator $g \in G$ at time period t, [kW]
$\hat{P}_t^b$	EMS set-point given to the ESS $b \in B$ at time period t, [kW]

Microgrid components:

P_t^b	Output power of the $b \in B$ at time period t, [kW]
$P_t^{b\leftarrow}$	Charge power of $b \in B$ at time period t, [kW]
$P_t^{b\rightarrow}$	Discharge power of $b \in B$ at time period t, [kW]
curt_t	Total curtailment applied to the power generation at time period t, [kW]
P_t^{res}	Power utilization of each RES $res \in RES$ at time period t, [kW]
S_t^b	Energy available inside each ESS $b \in B$ at time period t, [kWh]
nse_t	Not supplied energy at time period t, [kWh]

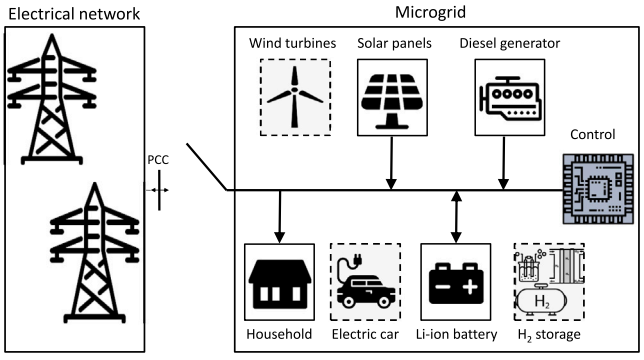


Fig. 1. Microgrid scheme.

balanced operation of the microgrid. It includes decentralized control strategies, such as droop control for distributed generation units and reactive power control for maintaining voltage stability. Secondary control focuses on restoring frequency and voltage deviations that occur after the primary control actions. It involves centralized or distributed communication-based control schemes aiming at keeping energy balance and optimizing the power flow within the microgrid. Finally, tertiary control, also known as EMS, is responsible for the economical and efficient operation of the microgrid, [5,6].

EMS design can be very challenging as it requires finding not only feasible but also optimal decisions for scheduling and dispatching the available generation resources, carrying out load management, if possible, and managing ESSs. EMSs should also coordinate with the main grid, managing energy imports and exports based on grid requirements and market conditions.

Literature shows that EMSs for isolated microgrids can be addressed through various optimization techniques and artificial intelligence methods [7,8]. Classical optimization techniques, such as Mixed Integer Linear Programming (MILP) [9,10], or Mixed Integer Quadratic Programming (MIQP) [11], offer deterministic or stochastic approaches to find optimal solutions. On the other hand, metaheuristic algorithms such as Particle Swarm Optimization (PSO), [12], explore the search space more flexibly to identify near-optimal solutions. Vilaisarn et al. [13] apply a combined approach where samples of the inner-level problem of a bilevel optimization are learned by Deep Learning (DL). Artificial intelligence methods, particularly DRL, have gained interest due to their ability to learn optimal control policies by interacting with the environment, making them suitable for managing complex, dynamic, and uncertain scenarios, [14–17]. In addition, as noted by Sutton and Barto in their seminal book on Reinforcement Learning (RL) [18], the trial-and-error nature of the technique lends itself well

to the development of EMS controllers that can then be applied to a wide range of microgrids without requiring tailor-made models or advanced computational capabilities. This characteristic could be extremely valuable in enabling the deployment and scaling up of the microgrid concept within future power systems.

From the modeling point of view, EMSs for microgrids face a significant challenge when it comes to accurately representing battery losses, among others, as they are nonlinear in nature. Linear modeling for optimization struggles to adequately incorporate these losses, where a common approach is to assume fixed efficiencies for the charge and discharge processes. As a result, there may be significant deviations between the expected and actual behavior in microgrids, thereby giving suboptimal results. This underscores the need for more sophisticated modeling techniques, although increasing the computational burden of the solution would be undesirable.

Theoretically, DRL techniques are able to address nonlinear dynamics in the environment because neural networks can capture nonlinearities. Several works utilize DRL or Dynamic Programming (DP)-based methods for the EMS problem [19–28], but few use a highly nonlinear model of the microgrid [22,24,25,28]. Actually, none of them analyzes the impact of a simplified linear model in the nonlinear nature, particularly on batteries, which have an important role when using RES. [19] utilizes a Deep Q-Network (DQN) for the EMS of a residential microgrid with a quadratic diesel cost. [20] proposes a TD3 for the EMS of a residential and a Low-Voltage Microgrid Benchmark (CIGRE) – as the proposed in this paper – with a quadratic diesel cost. [21] models a grid-connected microgrid with linear dynamics. Its authors compare several DRL algorithms over a time span of up to 168 h. [22] includes a nonlinear model of a thermostatically controlled load in its EMS, although the battery dynamics are linear. The span optimized is 10 days with an episode horizon of 24 h in each run, starting from the first hour of each day and selecting the day randomly. They propose to enhance the DRL algorithms used with a prioritized Experience Replay (ER) memory. [23] includes prices in its grid-connected microgrid model, includes future information, prices. The Proximal Policy Optimization (PPO) method proposed uses reward shaping but do not assess its benefits. [24] models a quadratic cost of the diesel, linear losses of the battery and nonlinearities in the power flow. The DRL algorithm used is MuZero. [25] uses the same model as [24] but with a custom DQN-based algorithm. [26] uses a similar model as ours, but with a linear model of the battery. Additionally uses an hydrogen electrolyzer/Fuel Cell (FC). [27] models a quadratic cost of the diesel. The algorithm used is not RL but a DP approach, that needs a model based on Markov Decision Process (MDP) as in RL. [28] models the nonlinear dynamics of the battery and a quadratic diesel cost. They do not uses a DRL, instead they use an Approximated Dynamic Programming (ADP) approach.

This paper introduces a novel approach to energy management in microgrids, utilizing DRL techniques. The main innovation lies in applying DRL, accounting for the nonlinear dynamic equations of battery losses in the microgrid system. This approach provides an approximated nonlinear model of real battery behavior, keeping the nonlinear complexity, which is crucial for the effective performance of energy management systems in real-world scenarios. Building upon previous works [19,20], in this study we have chosen the TD3 algorithm among others studied, due to its stability and capability of making decisions in the continuous domain [29].

The main contributions of this paper can be summarized as follows:

- First of all, a microgrid model that includes the nonlinear behavior of Lithium-Ion (Li-ion) batteries is proposed for the training of the TD3 algorithm. This model extends the POMDP of the microgrid, developed previously [19,20], with the nonlinear equations of the battery losses.
- Secondly, a methodology to evaluate the extent to which the proposed algorithm improves with respect to EMSs models with linear battery losses. Additionally, this work proposes a novel consideration of the control system in the microgrid model, which improves the robustness of the proposed method and the reliability of the results.
- Thirdly, this method is extended to manage a microgrid of realistic size, ensuring that the proposed method is valid for both a residential microgrid and a low-voltage distribution network extended into a microgrid.

This paper is organized as follows: Details about issues regarding battery modeling are summarized in Section 2. Section 3 describes the MDP model used for the EMS problem. Section 4 describes the control strategy underneath the EMS as part of the MDP model and its implications in the optimization process. The study case and results can be found in Section 5. Finally, conclusions are summarized in Section 6.

2. Battery operation model

Batteries play a pivotal role in electrical energy storage due to their ability to store and release energy in a highly controllable manner. The intermittent and fluctuating nature of PV and WT systems require the deployment of efficient ESSs, and in this context, batteries, particularly Li-ion variants, have emerged as a promising solution to address these challenges. A battery is a highly complex system, and its modeling should be tailored to the specific application for which it is intended in order to ensure that the chosen model adequately represents battery behavior and performance characteristics relevant to the desired use case, balancing accuracy and computational efficiency. In [30], the authors present a review of the literature on different approaches to model Li-ion batteries. Broadly speaking, one can distinguish between Electrochemical Models (EMs) and electrical Equivalent Circuit Models (ECMs). For an EMS in a microgrid, it is sufficient to use an ECM to represent the voltage, the State of Charge (SoC), and the power capabilities of the battery rather than using a detailed EM. Hence, an ECM has been used in the work reported here.

2.1. Equivalent circuit model

The Shepherd model [31] describes the output voltage of a battery in the discharge process as:

$$V_{batt}(t) = E_o - R \cdot i(t) - K\frac{Q(t)}{Q(t) - it(t)}i(t) + A \cdot e^{-B \cdot \frac{it(t)}{Q(t)}} \tag{1}$$

where V_{batt} is the voltage of the battery, E_o is the constant potential of the cell, R is the internal resistance, i is the current withdrawn from the battery, K is the polarization coefficient, Q is the amount of available charge, it is the total electrical charge extracted from the battery at time t measured from the moment that the discharge started ($it = \int i dt$), and A and B are constants to model the initial exponential drop expressed in the last term of (1). The values of E_o, K, Q, R, A, and B must be determined empirically.

In [32], the authors start from an equation similar to (1) in which they introduce some improvements that consider not only a constant discharge current but also the case of a variable charging or discharging current. The authors of Nguyen and Crow [33] take the equations from Tremblay and Dessaint [32] and derive simplified expressions of the voltage drop in the battery that is used to estimate the discharge/charge losses. Shuai et al. [28] refine previous expressions to obtain the following nonlinear equations of battery losses that will be used in this paper:

$$\begin{aligned} P^{b\leftarrow}_{loss} = & \frac{10^3(R_{in} + \frac{K}{1.1 - SOC})}{V_r^2}(P^{b\leftarrow})^2 \\ & + \frac{10^3 S^b_{\max} K(1 - SOC)}{SOC \cdot V_r^2} P^{b\leftarrow} \end{aligned} \tag{2}$$

$$P_{loss}^{b\rightarrow} = \frac{10^3(R_{in} + \frac{K}{SOC})}{V_r^2}(P^{b\rightarrow})^2 + (\frac{10^3 S_{\max}^b K(1-SOC)}{SOC \cdot V_r^2})P^{b\rightarrow} \quad (3)$$

where

- $P^{b\leftarrow}$: power consumed by the battery in [kW]
- $P_{loss}^{b\leftarrow}$: power losses while charging [kW]
- $P^{b\rightarrow}$: power generated by the battery [kW]
- $P_{loss}^{b\rightarrow}$: power losses while discharging [kW]
- R_{in}: internal resistance [Ohm]
- SOC: state of charge expressed in terms of the estimated stored energy [%]
- V_r: nominal voltage rate of the battery [V]
- $S_{\max}^b$: nominal capacity rate of the battery [kWh]

As explained in [34], the second terms of expressions (2) and (3) could be dismissed in order to obtain a more accurate model of the losses. This would result in the following expressions of the losses:

$$p_{loss}^{\leftarrow} = 10^3(R + \frac{K}{1.1 - SOC})(\frac{p_t^{\leftarrow}}{V_r})^2 \quad (4)$$

$$p_{loss}^{\rightarrow} = 10^3(R + \frac{K}{SOC})(\frac{p_t^{\rightarrow}}{V_r})^2 \quad (5)$$

Using the expressions of the losses, the energy stored in the battery S^b changes according to (6) and (7) for the charge and discharge, respectively.

$$\frac{dS^b}{dt} = P^{b\leftarrow} - P_{loss}^{b\leftarrow} \quad (6)$$

$$\frac{dS^b}{dt} = -P^{b\rightarrow} - P_{loss}^{b\rightarrow} \quad (7)$$

Energy flows in a battery are depicted in Fig. 2.

Discrete-time versions of (6) and (7) for the charge and discharge process, respectively, can be written as:

$$S_{t+1}^b - S_t^b = [P_t^{b\leftarrow} - P_{loss}^{b\leftarrow}(P_t^{b\leftarrow}, S_t^b)]\Delta t \quad (8)$$

$$S_{t+1}^b - S_t^b = [-P_t^{b\rightarrow} - P_{loss}^{b\rightarrow}(P_t^{b\rightarrow}, S_t^b)]\Delta t \quad (9)$$

These discrete-time equations are more convenient than the continuous-time ones in the context of this paper while being accurate enough. Therefore, from now on, magnitudes that express power as P will represent the average power in the discrete-time interval between t and $t + \Delta t$, i.e.:

$$P_t \Delta t = \int_t^{t+\Delta t} P dt \quad (10)$$

Similarly, energy S at time t in the battery b will represent the energy at the beginning of the continuous interval between t and $t + \Delta t$.

3. EMS of a microgrid using DRL

Microgrids combine several components that interact with each other, sharing energy at controllable power levels. These are classified into power generation units, ESSs, and loads. In the DRL context, the EMS represents the agent that interacts with each one of the controllable components within the microgrid. This process is denoted as MDP.

3.1. Markov decision process

The microgrid management is modeled as a discrete-time process using the MDP notation, i.e., a set of states $\mathbb{S}$, a set of actions $\mathbb{A}$, a reward function $R : \mathbb{S} \times \mathbb{A} \times \mathbb{S} \rightarrow \mathbb{R}$ that assigns a reward value to each transition (s, a, s'), and a set of conditional transition probabilities $T(s'|s, a)$. Each of these elements is detailed below.

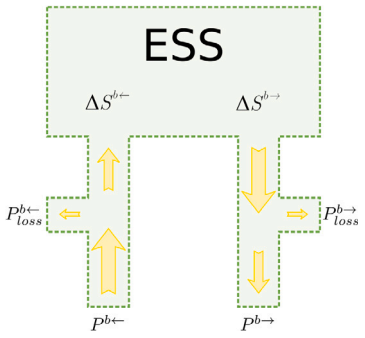


Fig. 2. Energy flow in a battery using the nomenclature defined in this paper.

3.1.1. State definition

The state gives full information about the problem to solve, satisfying the Markov Property [18,35]. In the EMS, the state is composed of exogenous and endogenous variables and may comprise the temporal dimension, i.e., information from the past. Exogenous variables correspond to the power dispatched by RESs and loads in the microgrid, whereas endogenous state variables quantify the energy stored in each of the ESSs.

Formally, the average active power in t for each RES $res \in Res$ in the microgrid is defined as P_t^{res}. Likewise, the demand for each load $l \in L$ over t is P_t^l. The energy stored inside each ESS $b \in B$ at the beginning of each period t is defined as S_t^b.

The domain of each variable is:

$$\begin{aligned} P_t^{res} &\in \mathbb{R}^+ && \forall res \in Res, \forall t \in T \\ P_t^l &\in \mathbb{R}^+ && \forall l \in L, \forall t \in T \\ S_t^b &\in (S_{\min}^b, S_{\max}^b) && \forall b \in B, \forall t \in T \end{aligned} \quad (11)$$

3.1.2. Action definition

In the MDP context, actions are set-points that the EMS sends to each component controller. These actions are defined as:

$$a_t = \{\hat{P}^a \mid a \in G \cup B\} \quad (12)$$

where $\hat{P}_t^g$, $g \in G$ is the active power generation set-point given for the controllable generator g and $\hat{P}_t^b$, $b \in B$ is the active power dispatch or consumption set-point for the ESS b.

These set-points sent to each controllable component may differ from the achievable values. Control systems could be forced to deviate from the set-point given in order to ensure the stability of the component. Therefore, for each set-point given $\hat{P}_t^a$, there is a consequent value P_t^a computed at the end of the period $[t, t + \Delta t]$ corresponding to the component $a \in G \cup B$. This behavior is formalized as a general function g:

$$P_t^a = g(\hat{P}_t^a, \ldots) \quad \forall a \in G \cup B, t \in T \quad (13)$$

and further detailed in Section 4.

Formally, the action domain, defined by its parts, is as follows:

$$\begin{aligned} \hat{P}_t^a &\in \mathbb{R} && \forall a \in G \cup B \\ P_t^g &\in \{0\} \cup [P_{\min}^g, P_{\max}^g] && \forall g \in G, \forall t \in T \\ P_t^b &\in [-P_{\max}^{b\leftarrow}, P_{\max}^{b\rightarrow}] && \forall b \in B, \forall t \in T \end{aligned} \quad (14)$$

where $P_{\min}^g$ and $P_{\max}^g$ are the minimum and maximum power values that the power generator can dispatch when it is on. Additionally, the generators can be turned off; thereby, the 0 value is considered in the action domain. $P_{\max}^{b\leftarrow}$ and $P_{\max}^{b\rightarrow}$ are the maximum power values when charging and discharging the battery, respectively.

3.1.3. Reward definition

The agent is designed to achieve a main goal and is guided toward this objective by reward signals, which the agent perceives immediately after making a decision. These signals indicate how well the agent's action is aligned with the goal. In the microgrid, where the agent is the EMS, this goal is to minimize the operational costs. Notice that these costs are negative rewards, encouraging the agent to reduce them.

Formally, the reward value r_t, which corresponds to the cost of energy produced from t up to $t + \Delta t$, is given by:

$$\begin{aligned} r_t &= R(s_t, a_t, s_{t+1}) \\ &= -\sum_{g}^{G} C^g(P_t^g) \cdot \Delta t - C^{\text{nse}} \cdot \text{nse}_t \end{aligned} \tag{15}$$

where C^g is the cost function of the generator that depends on the power dispatched P_t^g in the same period t. C^{nsp} is the cost for a unit of energy not supplied nse_t.

The cost of each generator is modeled as a quadratic function defined by its coefficients δ_i [27]:

$$C^g(P) = \begin{cases} \delta_2^g (P)^2 + \delta_1^g P + \delta_0^g & \text{if } P > 0 \\ 0 & \text{if } P = 0 \end{cases} \tag{16}$$

3.1.4. Transition definition

In MDPs, transitions link states and actions, and they can be represented as directional graphs where each node represents either a state or an action. Each transition is characterized by a probability of evolving from one system state to another in response to a particular control action. In the EMS example, these transitions represent the dynamics of the different components and the intermittency of renewable sources. The dynamics of the microgrid components have been defined by the equations below. In contrast, the intermittency of renewable sources has not been modeled by equations but by using time-series data as in [26].

The energy inside battery b satisfies the following energy balance equation:

$$S_{t+1}^b = S_t^b + \left[P_t^{b\leftarrow} \cdot \eta^b(P_t^{b\leftarrow}, S_t^b) - P_t^{b\rightarrow} \frac{1}{\zeta^b(P_t^{b\rightarrow}, S_t^b)} \right] \Delta t \tag{17}$$

where $P_t^{b\leftarrow}$ and $P_t^{b\rightarrow}$ are the charge and discharge battery power, and η^b and ζ^b are the corresponding nonlinear efficiency values for the charge and discharge processes, respectively. In particular, these efficiencies are defined as follows:

$$\eta^b(P^{b\leftarrow}, S^b) = \frac{P^{b\leftarrow} - P_{loss}^{b\leftarrow}(P^{b\leftarrow}, S^b)}{P^{b\leftarrow}} \tag{18}$$

$$\zeta^b(P^{b\rightarrow}, S^b) = \frac{P^{b\rightarrow}}{P^{b\rightarrow} + P_{loss}^{b\rightarrow}(P^{b\rightarrow}, S^b)} \tag{19}$$

for all $b \in B$, using the Eqs. (2) and (3). In batteries, the charge and discharge processes cannot happen at the same time, that formally can be modeled with the following constraint:

$$P_t^{b\leftarrow} \perp P_t^{b\rightarrow} \quad \forall b \in B, t \in T \tag{20}$$

which represents that these two variables are orthogonal w.r.t. each other, i.e., at least one variable has to be 0. Hence, P_t^b is defined from $P_t^{b\leftarrow}$ and $P_t^{b\rightarrow}$ by the equation:

$$P_t^b = P_t^{b\rightarrow} - P_t^{b\leftarrow} \quad \forall b \in B, t \in T \tag{21}$$

Additionally, the load balance equation must be satisfied in every period t:

$$(D_t + \text{curt}_t)\Delta t = (P_t^{Res} + P_t^G + P_t^B)\Delta t + \text{nse}_t \ \forall t \in T \tag{22}$$

where curt_t is the power curtailment in the microgrid during the period t and D_t, P_t^G, P_t^{Res}, and P_t^B are the totals in t for each component kind, and mathematically defined as:

$$\text{(total load)} \qquad D_t = \sum_{l}^{L} P_t^l \ \forall t \in T \tag{23}$$

$$\text{(total fuel-based gen.)} \qquad P_t^G = \sum_{g \in G} P_t^g \ \forall t \in T \tag{24}$$

$$\text{(total RES gen.)} \qquad P_t^{Res} = \sum_{res \in Res} P_t^{res} \ \forall t \in T \tag{25}$$

$$\text{(total ESS gen./load)} \qquad P_t^B = \sum_{b \in B} P_t^b \ \forall t \in T \tag{26}$$

3.2. Partially observable Markov decision process

Since the microgrid state is not fully observable [15], the problem must be modeled using a POMDP, which is a generalization of the MDP. In addition to the elements of MDPs, POMDPs include a set of observations Ω and a set of conditional observation probabilities O [36].

To increase the likelihood over the information input, a set of h consecutive observations are stacked:

$$h_t = (o_{t-h+1}, \dots, o_{t-1}, o_t) \ \forall t \in T \tag{27}$$

In the EMS, the observations for each period t are defined as,

$$o_t = (\mathbf{P}_{t-1}^{Res}, \mathbf{P}_{t-1}^{L}, \mathbf{S}_t^B) \ \forall t \in T \tag{28}$$

where

$$\begin{aligned} \mathbf{P}^{Res} &= \{P^{res} \mid res \in Res\} \\ \mathbf{P}^{L} &= \{P^{l} \mid l \in L\} \\ \mathbf{P}^{B} &= \{P^{b} \mid b \in B\} \end{aligned} \tag{29}$$

Notice that the exogenous variables P_t^{res} and P_t^l are unknown at the beginning of any period t, and therefore, $t - 1$ information is used instead.

3.3. DRL using TD3

DRL are based on learning by trial-and-error, i.e., by the interaction between the agent and the environment. From this interaction, the agent perceives rewards that help to improve its next actions [18].

In this work, the EMS is optimized by the TD3, which extends the well-known DQN technique, famous for surpassing the human performance in 49 Atari Games [37]. TD3 mainly adds the ability to deal with actions in a continuous domain [29,38], which is a critical feature in EMSs [20].

Previous results with the TD3 [20] motivated the authors of this paper to apply the same base algorithm. However, in that work, battery dynamics were modeled by linear equations. Given the fact that DRL can capture nonlinear relationships, this paper extends the previous microgrid model with nonlinear equations for battery dynamics to have a more realistic scenario.

4. Including the control system in the model

Practical approaches for implementing a controller include, beyond the EMS, an underlying real-time control system that will eventually guarantee the balance between generation and loads. Due to the battery operation flexibility, this is traditionally done by letting the control system directly manage the battery and excluding the battery action as a decision for the RL-based EMS [26], i.e., the battery is not given any reference by the EMS. This approach can be generalized by selecting any other component of the microgrid, instead of the battery, to take care of the balance in real-time. To avoid the complications when the balancing device saturates, this paper proposes using more than one component to take care of the balance.

4.1. Modeling the control system by using a priority list

Any EMS set-points may be ignored for the benefit of safety. When a component saturates, i.e., it cannot keep the balance without violating

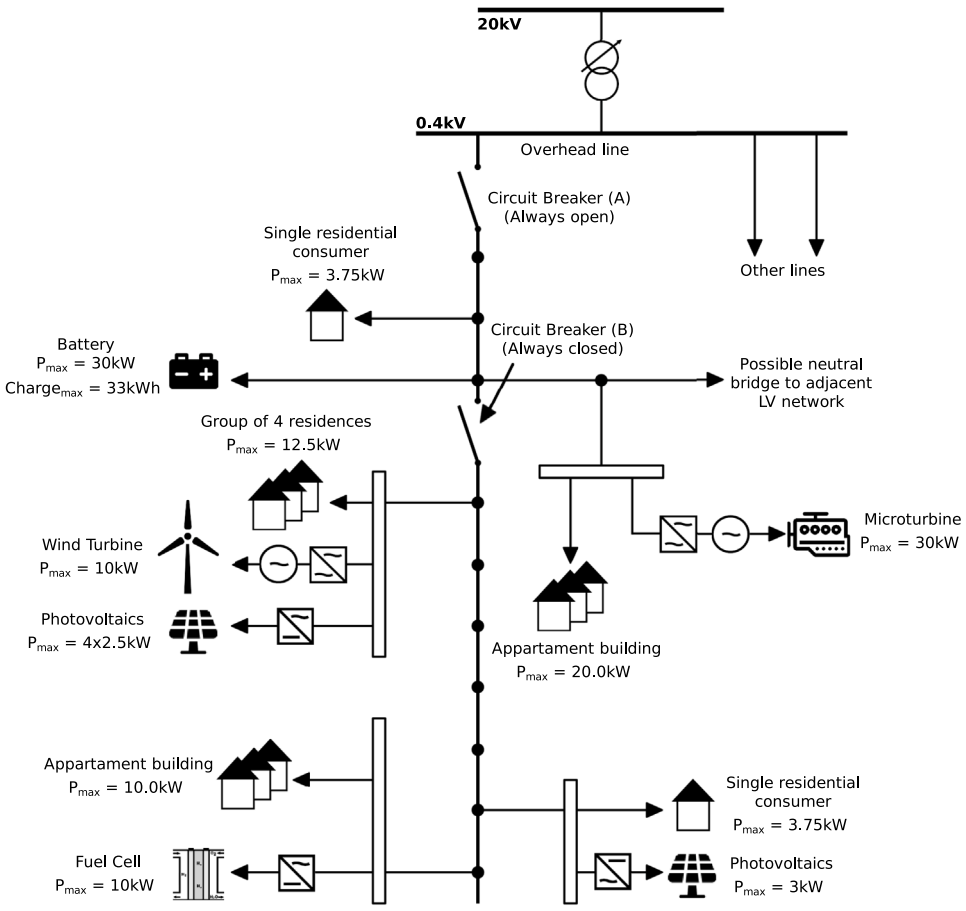


Fig. 3. Microgrid CIGRE.

its physical constraints, another component must replace the role of the saturated one. For this purpose, a priority list of components is defined to reinforce balance stability as much as possible. For example, let us consider the isolated microgrid of Fig. 1 using the components that are connected. When the battery is full, and the diesel unit has reached its power limit, if the demand is suddenly reduced, the microgrid will have an energy surplus that the battery cannot store. In this situation, the classical approach would spill the energy surplus. On the contrary, if there were a priority list of components, the diesel unit could be chosen to take care of the balance in this example, i.e., the control system would decrease the diesel power, thereby minimizing the operation costs. This behavior has been added in the RL environment-agent loop of the proposed POMDP in (13).

Although this approach protects the EMS and reduces the operational costs, it may also damage the trial-and-error learning process of the DRL agent since the error perceived after making a bad decision is reduced. Technically, the agent will perceive a more sparse reward, which does not benefit the learning process [39]. The strategy used in this paper during the experiments is to combine both approaches, i.e., to apply the classical approach during the agent's training and validate its performance with the second approach using the priority list. This combination takes the best of both worlds: it avoids slowing down the learning process while correctly assessing the agent performance on the more realistic microgrid operation.

In a real-time scenario, the chosen balancing device will need a primary controller to stabilize the frequency. A detailed model of this lower-level control is beyond the scope of this paper.

5. Performance comparison between linear vs. nonlinear Li-Ion battery models

This section highlights the advantages of using DRL methods for an EMS when considering Li-ion batteries. These methods employ Neural Networks (NNs) at their core, whose proficiently approximate nonlinear dynamics, as the Universal approximation theorem states. This ensures an enhanced adaptation to the complex behavior of the batteries, optimizing the EMS performance and reliability in real-world applications. The simulation experiments consider two isolated microgrids of different configuration sizes: a residential microgrid and the CIGRE [40].

The residential microgrid constitutes a typical demand pattern in a household with a PV, a Diesel Generator (Di-Gen), and a Li-ion battery, which is depicted in Fig. 1 considering only the components that are connected by lines in that figure, whereas the CIGRE constitutes a bigger microgrid, which is depicted in Fig. 3. The parameters of all elements in both case studies are shown in Table 1.

The datasets used in each microgrid for the demand, the PV and the WT, are composed of three hourly years. PV and demand datasets come from François-Lavet et al. [26], whereas the WT dataset comes from Renewables Ninja [41,42].

Given the limited data, it is necessary to assess the generalization capacity of the model [43,44] and a Machine Learning (ML) methodology has been used for this purpose. The dataset is divided into three distinct subsets: the training set, the validation set, and the test set. These subsets facilitate implementing a robust model validation

Table 1
Component specifications of the microgrid.

Element	Parameter	Resi.	CIGRE	Unit
Load (total)	$D_{\max}$	2.1	40.0	[kW]
PV (total)	$P^{pv}_{\max}$	6.0	13.0	[kW]
WT	$P^{pv}_{\max}$	6.0	10.0	[kW]
Di-Gen/Microturbine (MT)	$P^{d/mt}_{\max}$	1.0	30.0	[kW]
	$P^{d/mt}_{\min}$	0.1	3.0	[kW]
	$\delta^{d/mt}_0$	0.0157	0.4710	[€]
	$\delta^{d/mt}_1$	0.1080	0.1080	[€/kW]
	$\delta^{d/mt}_2$	0.3100	0.0103	[€/kW^2]
FC	$P^{fc}_{\max}$	–	10.0	[kW]
	$P^{fc}_{\min}$	–	0.0	[kW]
	δ^{fc}_0	–	0.0	[€]
	δ^{fc}_1	–	0.2	[€/kW]
	δ^{fc}_2	–	0.0	[€/kW^2]
Li-ion	$S_{\max}$	3.3	33.0	[kWESSh]
	$S_{\min}$	0.4	4.0	[kWESSh]
	S_0	0.4	4.0	[kWESSh]
	$P^{b\leftarrow}_{\max}$	3.0	30.0	[kW]
	$P^{b\leftarrow}_{\min}$	0.0	0.0	[kW]
	$P^{b\rightarrow}_{\max}$	3.3	33.0	[kW]
	$P^{b\rightarrow}_{\min}$	0.0	0.0	[kW]
	η-linear	0.9	0.9	[kWESSh/kWh]
	ζ-linear	0.9	0.9	[kWh/kWESSh]
	number of cells	1	10	[p.u.]
Li-ion individual cell (nonlinear)	internal resistance cons. (R_{in})	0.01	0.01	[Ω]
	nominal voltage (V)	51.8	51.8	[V]
	polarization constant (K)	0.06	0.06	[V/Ah] [Ω]
NSE	C^{nse}	1	10	[€/kWh]

Table 2
Yearly cost of each algorithm.

		Obj. F. ↔ Cost [€]		
Algorithm	Training	1st-Year	2nd-Year	3rd-Year
TD3	Linear	1665.46	1442.65	1545.39
TD3	Nonlinear	1628.42	1431.44	1518.12
Upper Bound	–	1569.67	1367.50	1427.66

process complemented by the early stopping technique to mitigate the risk of overfitting. Specifically, the dataset is divided uniformly using contiguous data; each subset represents a one-year span with hourly data points. More details are available in [20].

The following subsections analyze the results obtained from applying the TD3 method to two different battery models: a linear and a nonlinear model. First of all, the methodology to compare the approach with both battery models is detailed in Section 5.1. Secondly, the operation costs comparing both battery-loss models are discussed in Section 5.2 for each study case. Finally, in Section 5.3, the battery efficiency and energy losses are analyzed and compared between models.

5.1. Comparison methodology

The comparison made in this paper involves two POMDP approaches for the microgrid system, each characterized by a different energy loss model of the battery, i.e., using Eqs. (19) and (18) for the efficiencies in the nonlinear model, and constant values for the linear one. Consequently, these variations lead to differential behaviors in the microgrid system. In this sense, the TD3 trained using the linear model of the battery (TD3-L) and the same but using the nonlinear model of the battery (TD3-NL) are both evaluated in the microgrid using the nonlinear model of the battery, allowing a fair and reliable comparison since the same microgrid is used and that microgrid model is the closer to a real one. Fig. 4 depicts the comparison methodology.

Regarding the control system detailed in Section 4, the priority list to select the component responsible for taking care of the balance should be predefined for the evaluation stage. In this paper, for the residential microgrid, the battery comes first, and the diesel group

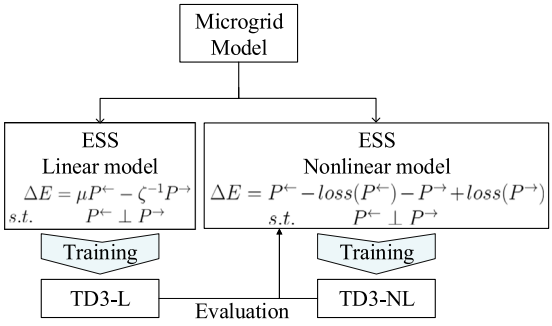


Fig. 4. Train-eval methodology.

second, while for the CIGRE microgrid, the battery comes first, the microturbine second, and the fuel cell comes third. This order has been chosen with the aim of using the most flexible and cheapest first.

5.2. Operation costs

5.2.1. Residential microgrid

Table 2 shows the total operational cost of operating the microgrid over three years. Additionally, the table includes the results from using a MIQP model, solved with the Gurobi solver [45] in a fully informed deterministic scenario. This last method is combined with a rolling horizon of 24 hours [46,47], and serves as an approximated upper bound to elucidate the goodness of the TD3. Notice that the MIQP cannot handle the nonlinear equation (17), and the following linear equation is used instead:

$$S^b_{t+1} = S^b_t + \eta^b P^{b\leftarrow} - P^{b\rightarrow}_t \frac{1}{\zeta^b} \; \forall b \in B, t \in T \tag{30}$$

where η and ζ are constants, which both take the value of 0.9.

The results in Table 2 show that the TD3-NL outperforms the TD3-L when both are tested with a nonlinear battery model. Compared with the TD3-L, the TD3-NL costs are reduced by 37.04€, 11.21€ and 27.27€ in each consecutive year. In addition, both TD3 configurations perform quite efficiently in the third year (the test dataset) when

Table 3
Yearly cost of each algorithm (CIGRE).

		Obj. F. ↔ Cost [€]		
Algorithm	Training	1st-Year	2nd-Year	3rd-Year
TD3	Linear	34 044.10	31 293.51	33 793.13
TD3	Nonlinear	33 279.29	30 633.29	33 158.29

Table 4
Energy losses of each algorithm and the difference between them (Residential).

	Energy Losses [kWh]		
Model	1st-Year	2nd-Year	3rd-Year
Linear	473.1695	502.5786	510.7831
Nonlinear	298.8784	285.2664	300.2103
Difference	174.2911	217.3122	210.5728

compared with the first and second years (the training and evaluation datasets). For instance, as the percentage relative error (RE) (see formula in (31)) with respect to the Upper Bound (the theoretical best), TD3-L has an RE of 8.25% in the third year versus 6.10% and 5.50% in the first and second year respectively. Similarly, TD3-NL has 6.34% in the third year versus the 3.74% and 4.68% in the first two. These results imply savings of 2.31% the first year, 0.82% the second, and 1.91% the third when using nonlinear battery dynamics in the model.

$$RE = \frac{|\text{measure} - \text{theoretical}|}{\text{theoretical}} \cdot 100\% \tag{31}$$

5.2.2. CIGRE microgrid

The CIGRE microgrid's results mirror the residential case on a proportional scale. Table 3 shows savings of 764.81€, 660.22€ and 634.84€ in each consecutive year when using the TD3-NL. These savings correspond to the 2.25%, 2.11%, and 1.88% with respect to the TD3-L costs using the formula in (31).

The TD3-NL performs better when operating the larger microgrid, indicating that the algorithm can handle different-size problems seamlessly. An extended analysis can be found in Appendix.

5.3. Battery efficiency and energy losses

Beyond the total costs, this paper analyzes additional metrics related to battery management, such as battery efficiency during charge and discharge processes, and the energy loss in the battery after each charge/discharge operation. Efficiency and energy loss metrics are strongly related to energy utilization (by Eqs. (6) and (7)) and can be used to analyze battery management performance. Efficiency helps visualize the operation patterns, whereas energy loss helps to quantify these patterns.

5.3.1. Residential microgrid

Regarding the residential case study, Figs. 5 and 6 show histograms of the battery efficiency during the discharge and charge processes, where each bar of the histogram corresponds to the number of hours the battery was operated with a particular efficiency (see formulas (18) and (19)). These histograms are also combined with the Kernel Density Estimation (KDE) curve and its average value (the vertical dashed line), and include both TD3-NL and TD3-L results.

For the discharge process, the TD3-NL average efficiency is 0.9055, whereas that of the TD3-L is 0.8508 (an improvement of 6.4%); meanwhile, during the charge process, the TD3-NL achieves an average efficiency of 0.8092 whereas that of the TD3-L is 0.6822 (an improvement of 18.6%). These experiments indicate that the model approach not only reduces operational costs but increases battery efficiency as a consequence.

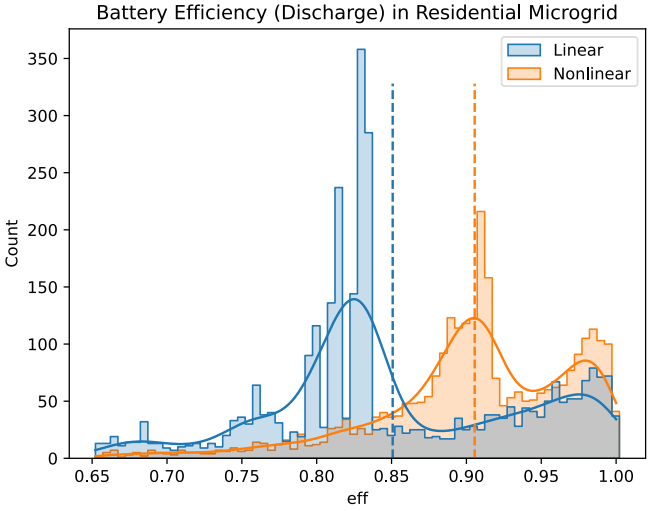


Fig. 5. Discharge efficiency comparison of TD3 trained using the linear and nonlinear battery model in a residential microgrid for the 3rd year.

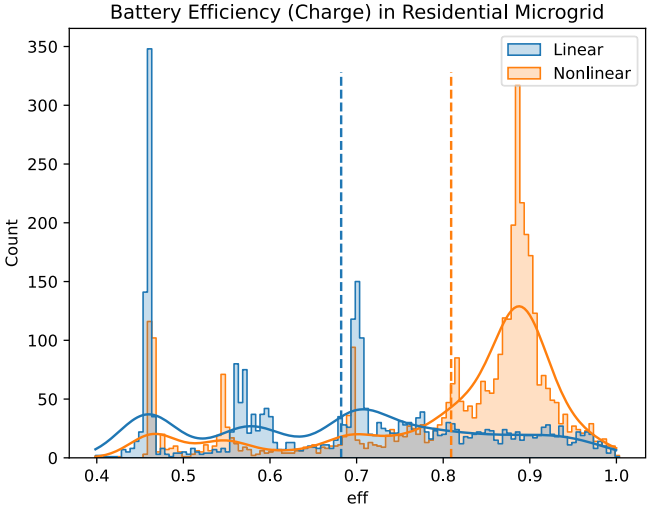


Fig. 6. Charge efficiency comparison of TD3 trained using the linear and nonlinear battery model in a residential microgrid for the 3rd year.

Fig. 7 shows a 3D scatter plot with the discharge efficiencies for the TD3-L(a) and TD3-NL(b). Similarly, Fig. 8 shows the same for the charge process. In both figures, the axes in the base represent the power and the SoC, and the vertical axis the efficiency, which also uses a color gradient to help visualize them (lighter means higher efficiency)

In Fig. 7(a), dots are sparse in the center of the area and more populated in the edges, whereas in (b) they are clustered in the high-efficiency area and in the low-power situations. Fig. 8 shows similar differences but more prominently since the charge process can reach very low efficiencies. These figures make visible the large change in the operation patterns of the battery.

Regarding battery energy losses, Table 4 shows that the consideration of the nonlinear battery model can drive the EMS to reduce the losses substantially. In particular, the TD3-L energy loss percentage over the total energy stored in the battery is 34.59%, 34.30%, 34.87% in each one of the three years, whereas the TD3-NL reduces it down to 24.17%, 22.96%, and 23.36%.

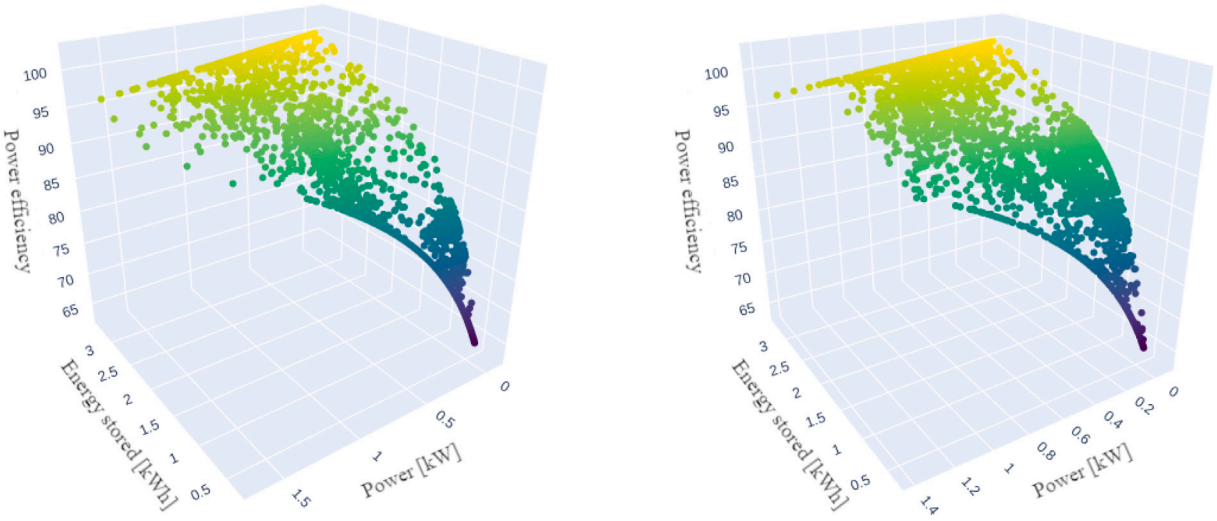


Fig. 7. 3D discharge efficiency for the TD3 in residential microgrid.

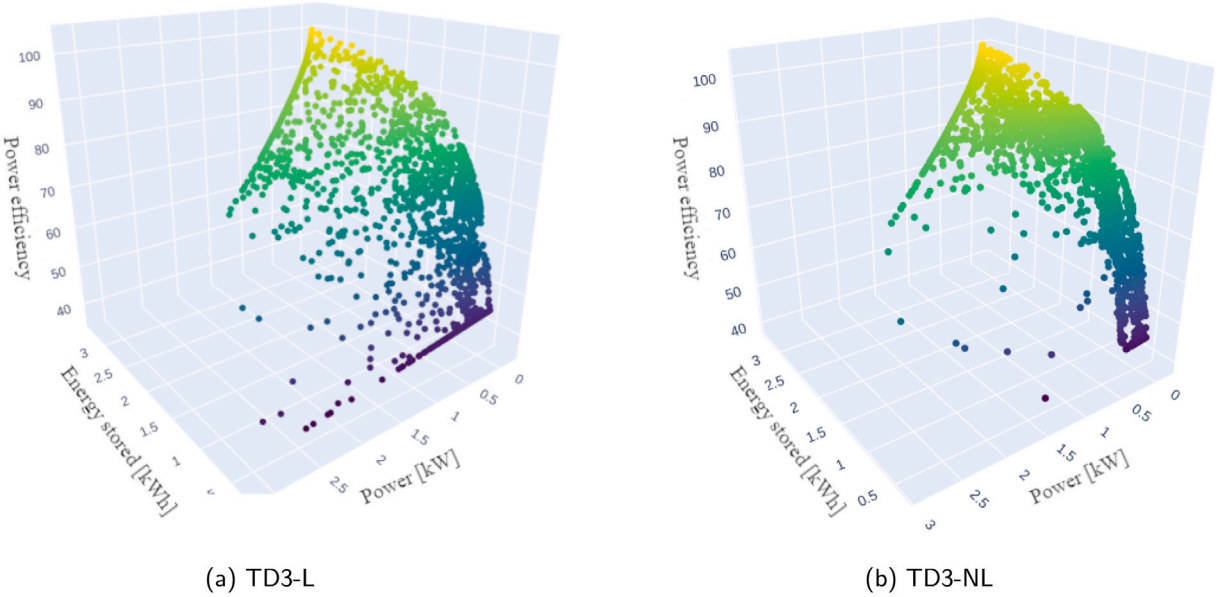


Fig. 8. 3D charge efficiency for the TD3 in residential microgrid.

5.3.2. CIGRE microgrid

Figs. 9 and 10 show the histograms corresponding to the CIGRE case study. KDE and the average value are depicted as in Figs. 5 and 6. During discharge, TD3-L achieves an average efficiency value of 0.9170 whereas TD3-NL achieves an average of 0.9421 (i.e., a 2.7% improvement). During charge, TD3-L achieves an average of 0.8246 whereas TD3-NL achieves an average of 0.9054 (i.e., a 9.8% improvement).

Regarding the energy losses of the battery in the CIGRE case, TD3-L obtains 26.67%, 26.46% and 26.65% of energy losses over the total energy stored in the battery, for each one of the three years whereas TD3-NL reduces it to 15.72%, 15.57% and 15.60%. The total energy losses by year are displayed in Table 5.

Table 5
Energy losses of each algorithm and the difference between both (CIGRE).

	Energy Losses [kWh]		
Model	1st-Year	2nd-Year	3rd-Year
Linear	6043.8106	5917.9887	6050.7169
Nonlinear	2585.0424	2524.1663	2561.6490
(Difference)	3458.7681	3393.8224	3489.0678

In conclusion, the results from the CIGRE case study are proportionally similar to the residential case study but with an absolutely huge difference given the scale between both.

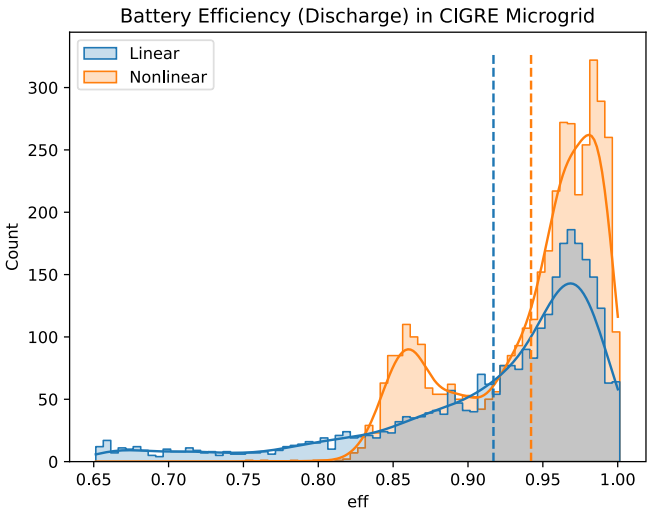


Fig. 9. Discharge efficiency comparison of TD3 trained using the linear and nonlinear battery model in CIGRE microgrid (3rd year).

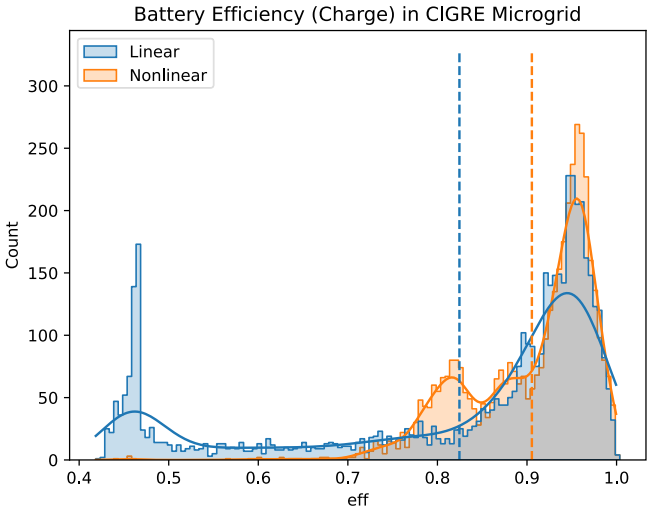


Fig. 10. Charge efficiency comparison of TD3 trained using the linear and nonlinear battery model in CIGRE microgrid (3rd year).

6. Conclusion

This paper introduces an application of the TD3 algorithm to an EMS in a microgrid and compares its performance using a linear (TD3-L) and a nonlinear (TD3-NL) Li-ion battery-loss model.

The research addresses several critical questions:

1. To what extent can the DRL-based algorithm comprehend and adapt to the inclusion of nonlinear battery loss dynamics, especially considering its limited information?
2. What is the behavior of the TD3 when integrated with a real-time control system that adjusts EMS decisions?
3. How effectively is the battery managed by the algorithm?

Our experiments provide satisfactory answers to these queries. The TD3 algorithm demonstrates a capacity to synergize with the control system, yielding near-optimal results in a highly uncertain environment. Moreover, the inclusion of nonlinear dynamics helps the algorithm to obtain even better results due to its ability to discern and leverage these nonlinear dynamics.

The findings reveal that incorporating nonlinear battery losses can result in approximately 2% savings in total microgrid operational costs, without comprising computational performance. Furthermore, the TD3-NL model significantly enhances battery efficiency, leading to around 50% in energy losses compared with the TD3-L. This translates to a savings of about 10% in energy losses over the total energy utilized through the battery.

An important observation is that the modeling efforts and computational resources required by the learning algorithm are similar in handling linear or nonlinear equations. This stands out as a major benefit over other optimization techniques.

CRediT authorship contribution statement

David Domínguez-Barbero: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Javier García-González:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Miguel Á. Sanz-Bobi:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Aurelio García-Cerrada:** Writing – review & editing, Validation, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

I have reference the links to the data in the references inside the manuscript.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT in order to improve readability and language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

This research has been funded by the Strategic Research Grants program of Comillas Pontifical University, and by Comunidad de Madrid, Spain and European Social Fund and the European Regional Fund (“ERDF a way of making Europe”) under the research programme "Microgrredes Inteligentes-Comunidad de Madrid, PROMINT-CM with reference number S2018/EMT-4366.

Appendix. Extended experimental results

This section analyzes several experiments on the CIGRE case in order to study the variability of the training and evaluation process using the DRL approach proposed. In particular, 6 learning trials with the linear model and another 6 with the nonlinear model of the battery. During this process, for each trial, the data from year 1 is used for training, and data from year 2 is used for validation (this is the same setup used for the results in Section 5). Once the models have been adjusted, they can be applied to any data series. Thus, they have been applied to year 1, year 2, and finally to year 3. It is worth mentioning that the time series data from last year 3 were not considered at any stage of the model adjustment, and therefore, it provides a good indicator of the model performance. The simulation results are depicted in Tables A.6 and A.7. Table A.6 shows the accumulated reward (in

Table A.6
Results from 6 trials using the linear model. Performance of the trained model (in euros) for each year.

	Obj. F. ↔ Cost [€]		
Linear	1st-Year	2nd-Year	3rd-Year
Trial 1	33138.86	30462.88	32962.70
Trial 2	33113.54	30388.39	32988.47
Trial 3	33242.64	30526.44	33107.89
Trial 4	34115.57	31449.56	33831.69
Trial 5	33306.55	30584.21	33219.58
Trial 6	33402.46	30614.49	33235.40
Average	33386.30	30671.00	33224.29
Std	372.75	390.12	318.33

Table A.7
Results from 6 trials using the nonlinear model. Performance of the trained model (in euros) for each year.

	Obj. F. ↔ Cost [€]		
Nonlinear	1st-Year	2nd-Year	3rd-Year
Trial 1	32748.90	30096.43	32620.46
Trial 2	33326.87	30700.60	33090.92
Trial 3	32688.01	30119.19	32628.19
Trial 4	32749.10	30086.15	32617.68
Trial 5	32782.68	30125.02	32656.95
Trial 6	32816.07	30170.34	32705.12
Average	32851.94	30216.29	32719.89
Std	236.52	239.05	184.71

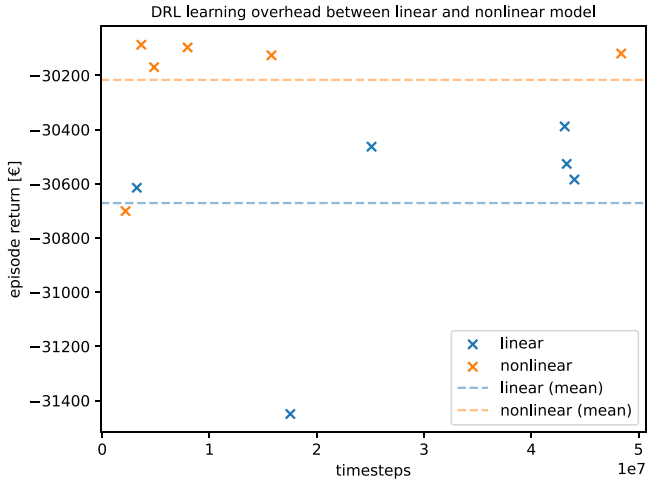


Fig. A.11. Computational burden required to train the model, measured in timesteps.

euros) for each of the 6 adjusted models with the linear losses applied for each one of the available years. Table A.7 shows the same results for the nonlinear case.

In each table, the variability of the results is due to the random initialization of the neural network and other random processes during the training, such as the exploration of the algorithm and the sampling from the ER memory. However, while this variability is a well-known characteristic that occurs every time a DRL model is adjusted, it is interesting to note that the obtained standard deviation is small. The absolute values of the coefficients of variation (i.e., the ratio between the standard deviation and the mean) are 1.12%, 1.27%, and 0.96% for the linear losses, and 0.72%, 0.79%, and 0.56% for the nonlinear losses.

So, we can say that the results are quite similar between different trials, concluding that the approach is stable, i.e., there is high confidence in a single trial to obtain an acceptable performance. Additionally, the table comparison shows that the nonlinear approach leads to better performance on average. Comparing the mean values for each year, the nonlinear model outperforms the linear one by 1.60%, 1.48%, and 1.52% for years 1, 2, and 3, respectively.

Moreover, these executions show the overhead in training the proposed approach. Fig. A.11 shows the computational burden, in timesteps, required to train the model and the performance of the best-chosen model. More details about the training process are in [20].

This figure shows that considering a nonlinear model of the Li-ion battery losses instead of a linear model does not imply an extra burden in the learning process. From the data observed, the average number of timesteps needed with the linear model is 29.38 million, and with the nonlinear model is 13.80 million. Besides, the same figure corroborates the outperforming of using the nonlinear model to train the proposed DRL approach.

References

[1] Renewable energy statistics. 2023, URL https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Renewable_energy_statistics. [Last access: 2023-12-09].

[2] Lasseter RH, Akhil AA, Marnay C, Stephens J, Dagle JE, Guttromson RT, et al. Integration of distributed energy resources: The CERTS MicroGrid concept. Report LBNL-50829, Berkeley, CA: Consortium for Electric Reliability Technology Solutions; 2003, p. 32.

[3] Akinyele D, Belikov J, Levron Y. Challenges of microgrids in remote communities: A STEEP model application. Energies 2018;11(2):432. http://dx.doi.org/10.3390/en11020432.

[4] Kaplan SM, Sissine F. Smart grid : Modernizing electric power transmission and distribution; energy independence, storage and security; energy independence and security act of 2007 (EISA); improving electrical grid efficiency, communication, reliability, and resiliency; integrating new and renewable energy sources. In: Government series, TheCapitol.Net, Inc.; 2009.

[5] Gao F, Kang R, Cao J, Yang T. Primary and secondary control in DC microgrids: A review. J Mod Power Syst Clean Energy 2019;7(2):227–42. http://dx.doi.org/10.1007/s40565-018-0466-5.

[6] Střelec M, Berka J. Microgrid energy management based on approximate dynamic programming. In: IEEE PES ISGT europe 2013. Lyngby, Denmark; 2013, p. 1–5. http://dx.doi.org/10.1109/ISGTEurope.2013.6695439.

[7] Zia MF, Elbouchikhi E, Benbouzid M. Microgrids energy management systems: A critical review on methods, solutions, and prospects. Appl Energy 2018;222:1033–55. http://dx.doi.org/10.1016/j.apenergy.2018.04.103.

[8] Lopez-Garcia TB, Coronado-Mendoza A, Domínguez-Navarro JA. Artificial neural networks in microgrids: A review. Eng Appl Artif Intell 2020;95:103894. http://dx.doi.org/10.1016/j.engappai.2020.103894.

[9] Sukumar S, Mokhlis H, Mekhilef S, Naidu K, Karimi M. Mix-mode energy management strategy and battery sizing for economic operation of grid-tied microgrid. Energy 2017;118:1322–33. http://dx.doi.org/10.1016/j.energy.2016.11.018.

[10] Amrollahi MH, Bathaee SMT. Techno-economic optimization of hybrid photovoltaic/wind generation together with energy storage system in a stand-alone micro-grid subjected to demand response. Appl Energy 2017;202:66–77. http://dx.doi.org/10.1016/j.apenergy.2017.05.116.

[11] Garcia-Torres F, Bordons C, Tobajas J, Real-Calvo R, Santiago I, Grieu S. Stochastic optimization of microgrids with hybrid energy storage systems for grid flexibility services considering energy forecast uncertainties. IEEE Trans Power Syst 2021;36(6):5537–47. http://dx.doi.org/10.1109/TPWRS.2021.3071867.

[12] Alavi SA, Ahmadian A, Aliakbar-Golkar M. Optimal probabilistic energy management in a typical micro-grid based-on robust optimization and point estimate method. Energy Convers Manage 2015;95:314–25. http://dx.doi.org/10.1016/j.enconman.2015.02.042.

[13] Vilaisarn Y, Rodrigues YR, Abdelaziz MMA, Cros J. A deep learning based multiobjective optimization for the planning of resilience oriented microgrids in active distribution system. IEEE Access 2022;10:84330–64. http://dx.doi.org/10.1109/ACCESS.2022.3197194.

[14] Glavic M, Fonteneau R, Ernst D. Reinforcement learning for electric power system decision and control: Past considerations and perspectives. In: 20th IFAC world congress, IFAC-PapersOnLine In: 20th IFAC world congress, 2017;50(1).6918–27,

[15] Francois-Lavet V, Henderson P, Islam R, Bellemare MG, Pineau J. An introduction to deep reinforcement learning. Found Trends Mach Learn 2018;11:219–354. http://dx.doi.org/10.1561/2200000071.

[16] Yang T, Zhao L, Li W, Zomaya AY. Reinforcement learning in sustainable energy and electric systems: A survey. Annu Rev Control 2020;49:145–63.

[17] Yu L, Qin S, Zhang M, Shen C, Jiang T, Guan X. Deep reinforcement learning for smart building energy management: A survey. 2020, arXiv:200805074 [cs, eess].

[18] Sutton RS, Barto AG. Reinforcement learning: An introduction. 2nd ed.. The MIT Press; 2018.
[19] Domínguez-Barbero D, García-González J, Sanz-Bobi MA, Sánchez-Úbeda EF. Optimising a microgrid system by deep reinforcement learning techniques. Energies 2020;13(11). http://dx.doi.org/10.3390/en13112830.
[20] Domínguez-Barbero D, García-González J, Sanz-Bobi MÁ. Twin-delayed deep deterministic policy gradient algorithm for the energy management of microgrids. Eng Appl Artif Intell 2022;13(11). http://dx.doi.org/10.1016/j.engappai.2023.106693.
[21] Panda DK, Turner O, Das S, Abusara M. Prioritized experience replay based deep distributional reinforcement learning for battery operation in microgrids. J Clean Prod 2024;434:139947. http://dx.doi.org/10.1016/j.jclepro.2023.139947.
[22] Nakabi TA, Toivanen P. Deep reinforcement learning for energy management in a microgrid with flexible demand. Sustain Energy Grids Netw 2021;25:100413. http://dx.doi.org/10.1016/j.segan.2020.100413, URL https://www.sciencedirect.com/science/article/pii/S2352467720303441.
[23] Lee S, Seon J, Sun YG, Kim SH, Kyeong C, Kim DI, et al. Novel architecture of energy management systems based on deep reinforcement learning in microgrid. IEEE Trans Smart Grid 2024;15(2):1646–58. http://dx.doi.org/10.1109/TSG.2023.3317096, URL https://ieeexplore.ieee.org/document/10255281. Conference Name: IEEE Transactions on Smart Grid.
[24] Shuai H, He H. Online scheduling of a residential microgrid via Monte-Carlo tree search and a learned model. IEEE Trans Smart Grid 2021;12(2):1073–87.
[25] Shuai H, Li F, Pulgar-Painemal H, Xue Y. Branching dueling Q-network-based online scheduling of a microgrid with distributed energy storage systems. IEEE Trans Smart Grid 2021;12(6):5479–82. http://dx.doi.org/10.1109/TSG.2021.3103405, URL https://ieeexplore.ieee.org/document/9509287. Conference Name: IEEE Transactions on Smart Grid.
[26] François-Lavet V, Taralla D, Ernst D, Fonteneau R. Deep reinforcement learning solutions for energy microgrids management. In: European workshop on reinforcement learning'13. Pompeu Fabra University, Barcelona, Spain; 2016, Permalink: https://hdl.handle.net/2268/203831.
[27] Jasmin EA, Imthias Ahamed TP, Jagathy Raj VP. Reinforcement learning approaches to economic dispatch problem. Int J Electr Power Energy Syst 2011;33(4):836–45. http://dx.doi.org/10.1016/j.ijepes.2010.12.008.
[28] Shuai H, Fang J, Ai X, Wen J, He H. Optimal real-time operation strategy for microgrid: An ADP-based stochastic nonlinear optimization approach. IEEE Trans Sustain Energy 2019;10(2):931–42. http://dx.doi.org/10.1109/TSTE.2018.2855039.
[29] Fujimoto S, van Hoof H, Meger D. Addressing function approximation error in actor-critic methods. In: Proceedings of the 35th international conference on machine learning. Proceedings of machine learning research, vol. 80, PMLR; 2018, p. 1587–96.
[30] Fotouhi A, Auger DJ, Propp K, Longo S, Wild M. A review on electric vehicle battery modelling: From Lithium-ion toward Lithium–Sulphur. Renew Sustain Energy Rev 2016;56:1008–21. http://dx.doi.org/10.1016/j.rser.2015.12.009.
[31] Shepherd CM. Design of primary and secondary cells: II. An equation describing battery discharge. J Electrochem Soc 1965;112(7). http://dx.doi.org/10.1149/1.2423659.
[32] Tremblay O, Dessaint L-A. Experimental validation of a battery dynamic model for EV applications. World Electr Veh J 2009;3(2):289–98. http://dx.doi.org/10.3390/wevj3020289.
[33] Nguyen TA, Crow ML. Stochastic optimization of renewable-based microgrid operation incorporating battery operating cost. IEEE Trans Power Syst 2016;31(3):2289–96. http://dx.doi.org/10.1109/TPWRS.2015.2455491.
[34] García-González J, Guerrero S. Optimal management of a microgrid Li-Ion battery considering non-linear losses using the integer zig-zag formulation. In: Proceedings of power systems computation conference. Paris-Saclay, France; 2024 [in press].
[35] Howard RA. Dynamic programming and Markov processes. 2nd ed.. John Wiley; 2018.
[36] Lauri M, Hsu D, Pajarinen J. Partially observable Markov decision processes in robotics: A survey. IEEE Trans Robot 2023;39(1):21–40. http://dx.doi.org/10.1109/TRO.2022.3200138.
[37] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature 2015;518(7540):529–33. http://dx.doi.org/10.1038/nature14236.
[38] Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. (DDPG) Continuous control with deep reinforcement learning. 2015, http://dx.doi.org/10.48550/arXiv.1509.02971, arXiv:150902971 [cs, stat].
[39] Rengarajan D, Vaidya G, Sarvesh A, Kalathil D, Shakkottai S. Reinforcement learning with sparse rewards using guidance from offline demonstration. 2022, http://dx.doi.org/10.48550/arXiv.2202.04628, arXiv:2202.04628 [cs].
[40] Papathanassiou S, Hatziargyriou N, Strunz K, et al. A benchmark low voltage microgrid network. In: Proceedings of the CIGRE symposium: Power systems with dispersed generation. CIGRE; 2005, p. 1–8.
[41] Stefan Pfenninger IS. Renewables ninja. Version: 1.1, coord: Lat. 39.459 - lon. -2.173, dates: 2018-01-01 - 2020-12-31, dataset: Merra2, capacity: 10kW, height: 80 m, turbine: GE 1.5.sle. 2023, URL https://www.renewables.ninja/. [Last access: 2023-11-27].
[42] Staffell I, Pfenninger S. Using bias-corrected reanalysis to simulate current and future wind power output. Energy 2016;114:1224–39. http://dx.doi.org/10.1016/j.energy.2016.08.068.
[43] Peng XB, Andrychowicz M, Zaremba W, Abbeel P. Sim-to-real transfer of robotic control with dynamics randomization. In: 2018 IEEE international conference on robotics and automation (ICRA). Brisbane, QLD, Australia; 2018, p. 3803–10. http://dx.doi.org/10.1109/ICRA.2018.8460528.
[44] Güitta-López L, Boal J, López-López ÁJ. Learning more with the same effort: How randomization improves the robustness of a robotic deep reinforcement learning agent. Appl Intell 2023;53:14903–17. http://dx.doi.org/10.1007/s10489-022-04227-3.
[45] Gurobi Optimization L. Gurobi optimizer reference manual. 2023, URL https://www.gurobi.com. [Last access: 2023-11-27].
[46] Le KD, Day JT. Rolling horizon method: A new optimization technique for generation expansion studies. IEEE Trans Power Appar Syst 1982;PAS-101(9):3112–6. http://dx.doi.org/10.1109/TPAS.1982.317523.
[47] Sethi S, Sorger G. A theory of rolling horizon decision making. Ann Oper Res 1991;29(1):387–415. http://dx.doi.org/10.1007/BF02283607.