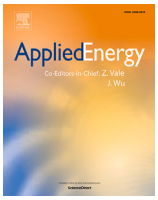
Applied Energy 400 (2025) 126465

Contents lists available at ScienceDirect

Applied Energy

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



Comparison of machine learning and MPC methods for control of home battery storage systems in distribution grids

Felicitas Mueller[a,*], Steven de Jongh[a], Claudio A. Cañizares[b], Thomas Leibfried[a], Kankar Bhattacharya[b]

[a] Institute of Electrical Energy Systems and High Voltage Engineering, Karlsruhe Institute of Technology (KIT), Engesserstraße 11, Karlsruhe, 76131, Baden-Württemberg, Germany

[b] Centre for Environmental & Information Technology, University of Waterloo, Ring Rd, Waterloo, N2L 3G1, Ontario, Canada

HIGHLIGHTS

- Applied Machine Learning and optimization for Home Energy Management Systems (HEMS).
- Comparison of rule-/model-based and model-free HEMS.
- Integrated Graph Neural Networks (GNNs) into HEMS control for the first time.
- Analyzed Imitation Learning trade-offs between computational demand and performance.
- Benchmarked HEMS methods using realistic low voltage grid and household and PV data.

ARTICLE INFO

Keywords:
Active distribution networks
Home energy management system
Imitation learning
Model predictive control
Neural networks
Optimal storage scheduling
Reinforcement learning

ABSTRACT

Control methods for Home Energy Management Systems implemented with traditional optimization techniques and state-of-the-art Machine Learning methods are presented and compared in this paper in the context of their impact on and interactions with Active Distribution Networks. Thus, model-based methods based on Model Predictive Control algorithms with different prediction qualities are first described and compared against model-free methods based on imitation learning and reinforcement learning. A practical, state-of-the-art, heuristic, rule-based controller is used as the baseline. An in-depth comparison is performed using metrics consisting of objective function values, grid constraint violations, and computational time. The results of applying these Home Energy Management Systems to a realistic German low voltage benchmark grid with 13 connected households, each containing solar generation, a battery storage system, and electrical loads are discussed. It is demonstrated that model-based and model-free methods can achieve improvements over typical rule-based methods, with varying performance in terms of objective function values and grid constraint violations depending on the forecasts, at the cost of higher computational complexity. Furthermore, model-free methods are shown to have in general low computational burden at higher objective function values with more grid constraint violations, with imitation-learning-based techniques proving to be the best compromise for practical applications.

1. Introduction

The ongoing energy transition is drastically impacting generation and consumption in power systems. This is particularly evident in Low Voltage (LV) grids due to the increasing penetration of Photovoltaic (PV) panels and Battery Energy Storage Systems (BESS). This leads to more weather-dependent loads with higher flexibility through the use of BESS, which integrate household appliances, HVAC systems, renewable energy sources, and BESS into a centralized optimization framework that aims at optimizing certain objectives [1–3]. In the existing literature, multiple model-based and model-free methods have been proposed with different levels of complexity for Home Energy Management Systems (HEMS), such as stochastic (e.g., Markov chains [4]), deterministic (e.g., perfect forecast Model Predictive Control (MPC) [5]), heuristic (e.g., Particle Swarm Optimization [6]), hierarchical multi-stage optimization (e.g.,

* Corresponding author.
Email address: felicitas.mueller@kit.edu (F. Mueller).

https://doi.org/10.1016/j.apenergy.2025.126465
Received 29 January 2025; Received in revised form 12 May 2025; Accepted 12 July 2025
Available online 2 August 2025

Glossary

A2C	Asynchronous Advantage Actor Critic
ARS	Augmented Random Search
BCF	Best-Case Forecaster
BESS	Battery Energy Storage System
CP	Control Policy
D-MPC	Decentralized Model Predictive Controller
DDPG	Deep Deterministic Policy Gradients
DQN	Deep Q-Network
DSO	Distribution System Operator
EV	Electric Vehicle
GDL	Geometric Deep Learning
GIL	Graph Imitation Learning
GNN	Graph Neural Network
HEMS	Home Energy Management System
HVAC	Heating, Ventilation and Air Conditioning
IL	Imitation Learning
L-MPC	Linearized Model Predictive Controller
LCPF	Linear Coupled Power Flow
LCQP	Linear Constrained Quadratic Programming
LV	Low Voltage
MARL	Multi-Agent Reinforcement Learning
MILP	Mixed-Integer Linear Programming
ML	Machine Learning
MLP	Multi-Layer Perceptron
MPC	Model Predictive Control
MSE	Mean Squared Error
MV	Medium Voltage
NL-MPC	Nonlinear Model Predictive Controller
NLP	Nonlinear Programming
NN	Neural Networks
NR	Newton Raphson
OPF	Optimal Power Flow
PCC	Point of Common Coupling
PPO	Proximal Policy Optimization
PV	Photovoltaics
RL	Reinforcement Learning
SAC	Soft-Actor-Critic
SC	Shortsighted Controller
SL	Supervised Learning
SOC	State Of Charge
TD3	Twin-Delayed DDPG
WCF	Worst-Case Forecaster

two-stage energy management [7]) and A.I.-based methods (e.g., deep neural networks [8]), demonstrating the various approaches that can be used to optimize schedules in HEMS. Some of the published methods have been trademarked or translated into patents (e.g., [9,10]). However, the widespread adoption of these complex methods has not yet occurred, which is evident from the fact that many of these methods have been available for over a decade, but have seen limited or no commercialization nor standard integration into mass production for newly built smart homes. Their implementation in practice has proven challenging due to the complexity of the proposed methods, which demonstrates that increasingly complex approaches are not really needed in practice. Therefore, this paper focuses on simpler, practical, and robust HEMS that can be quickly implemented with the existing technologies to reduce the overall complexity of controlling demand, which do not require a variety of interfaces and communication protocols, focusing more on the direct control of BESS rather than complex controls of appliances and HVAC systems. This is particularly relevant in practice for Distribution System Operators (DSOs), which are interested in the optimal interaction of actually deployed HEMS with the distribution grid.

The optimization of HEMS, which calculates the BESS schedules at the household level, plays a major role, with significant impacts on LV grids. However, depending on the objectives, HEMS can lead to local grid issues [11], highlighting the need for considering the different grid and household perspectives in practical HEMS applications. Thus, homeowners want to reduce electricity costs and increase resiliency, whereas grid operators are interested in minimizing the impact of HEMS on their grids, with the latter being the perspective taken in the current paper. These conflicting interests result in complex scheduling problems.

To tackle the aforementioned issues, various approaches have been presented in the literature, which can be categorized into two major categories [12,13], namely model-based methods, where the control algorithms are based on mathematical optimization models that model the physical system, and model-free methods, where the control algorithms are based on data-driven approximate models. Traditionally, the former approaches are often used for HEMS in practice [14], which allow for considering distribution grids [15] in the model formulation, resulting in complex optimization problems for DSOs based on Mixed-Integer Linear Programming (MILP) or Nonlinear Programming (NLP) methods. These techniques are complex due to the coupling of multiple time steps and the high number of flexible devices connected to the grid, posing a major challenge for DSO applications. To deal with this complexity, various linearization methods have been developed which simplify the load flow equations and thus reduce the complexity of the optimization problem [16]. Furthermore, MPC techniques are used to readily address load and weather uncertainty, resulting in model-based MPC approaches often being too complex and slow for practical applications. In this context, the present paper discusses practical MPC-based HEMS from a DSO perspective, which considers nonlinear and linearized load flow equations to quantify their impact on and interactions with LV grids.

In contrast to model-based HEMS approaches, model-free data-driven systems are learning controllers that can be either based on Supervised Learning (SL), where both inputs and outputs are known, or Reinforcement Learning (RL), where an optimal control policy is learned through interaction with a simulation model or real-world data sampling. SL and RL based HEMS range from Neural Networks (NNs) to random forest based regression methods. They show promising results [12,13], highlighting the value of data-driven approaches. Many different data-driven approaches have been proposed in the literature, but in-depth comparisons of these methods with heuristic and model-based methods are lacking. Therefore, this paper discusses the implementation of effective BESS scheduling model-based and model-free existing methods for DSOs to comprehensively compare them from a practical perspective.

A promising approach for the application of SL-based methods for optimal scheduling in electrical grids is based on Imitation Learning (IL), where the goal is to approximate the mapping from input data to output data created by model-based methods. It was shown in [17,18] that this approach can be used to derive fast approximators for single time-step OPFs based on linear regression and NNs, allowing for significant calculation speed-up. An extension to this approach was presented in [19], where a white-box HEMS MPC was for the first time imitated using NNs, showing an acceleration of the calculations by a factor of up to 50, while obtaining similar objective function values. While [19] investigates fast, approximative algorithms for scheduling such as NN, a linearly constrained MPC is used for training data generation. Hence, in the current paper, this consideration is extended by enhancing the proposed IL methods to nonlinear HEMS problems from a practical DSO perspective, including, for the first time, methods based on Graph Neural

Table 1
Overview of the state-of-the-art HEMS control algorithms.

Category	Algorithm	Comparison	Grid	Ref.
Model-based	Stochastic MPC	No controller	Yes	[28]
	Multi-level MPC	–	Yes	[29]
	MPC	–	No	[30]
Model-free	RL (Q-learning)	No controller	No	[31]
	IL MARL (DDPG)	MILP	No	[8]
	RL MARL (PPO)	MILP MPC	No	[32]
	RL (DQN)	Optimal control, Heuristics, IL	Yes	[33]
	RL (DQN)	No controller, MILP	No	[34]
	RL (DDPG)	Heuristics, Optimal control	No	[35]
	MARL (PPO)	Heuristics	No	[36]
	RL (A2C)	Manual controller, Rule-based controllers	No	[37]
	RL (DQN, DPG)	No controller	No	[38]
Hybrid	Learning-based MPC with Actor-Critic RL	–	No	[39]

Networks (GNNs), which allow taking the topology of the distribution grid into account.

While SL is dependent on an expert system, which provides the learning data for the trained approximators, RL can learn without this prerequisite. This eliminates the need to formulate any optimization problem for the scheduling task by DSOs by substituting the training data set with an Environment, on which the RL method is trained. RL-based approaches have been applied to various tasks in electrical grids [20–22], showing promising results but being limited to discrete action spaces due to the chosen Deep Q-learning approaches. This is a limiting factor for the use of RL for control of flexible devices in LV grids, as discussed in [23], where the action space is limited to six discrete actions for EV charging and ten discrete actions for BESS charging and discharging. This is inconsistent with the actual technical capabilities of BESS in practice, as these storage systems can operate continuously between their minimum and maximum limits, and thus, discrete control actions constrain their flexibility potential for the LV grid. Other methods that control continuous action spaces can also be found in the literature, with [24] focusing on transmission system economic dispatch and [25] on voltage control and loss minimization in distribution grids. However, these techniques have not been applied to HEMS problems, and are thus studied in the present paper based on [26]. Model-based and model-free methods differ in their computational demand, scalability, real-time capability, complexity, interpretability, accuracy, human modeling effort and robustness. A qualitative comparison of MPC and RL approaches for HEMS is given in [27], however, no quantitative analysis is provided, thus not allowing for a direct comparison.

Table 1 provides a comprehensive overview of the current state-of-the-art on HEMS control methods, classifying them as model-based, model-free, and hybrid approaches with model-based approaches corresponding to controllers that solve linear or nonlinear optimization problems using MPC techniques for a rolling-horizon. This table highlights the current trend towards model-free approaches based on Machine Learning (ML) techniques, mostly relying on RL [35,37] and Multi-Agent RL (MARL) [8,32,36] methods, often based on Q-learning and Deep Q-Networks (DQNs) [31,33,34,38], which are inherently limited by their discrete action space. Note that a comprehensive comparison between model-based and model-free algorithms is lacking, and that most existing studies focus on algorithms for a single HEMS without considering multiple buildings and their LV grid interconnection, limiting the conclusions that can be drawn from a DSO perspective.

A comparison of the different approaches for their application in flexibility scheduling is of great relevance for DSOs and homeowners in the context of practical HEMS applications. However, the existing literature lacks a proper, practical, quantitative comparison of these methods. Hence, this paper focuses on proposing, implementing and comparing such methods from a practical DSO perspective based on different key metrics. Based on the aforementioned discussions, the contributions of the paper can be summarized as follows:

- Developing model-based and model-free practical HEMS models for flexibility scheduling in LV distribution grids, considering both the interests of the DSO and the homeowners, while taking into account practical and realistic conditions for the different stakeholders.
- Applying for the first time GNNs to the HEMS nonlinear DSO problem, as well as studying existing RL and IL techniques with continuous action spaces for comparison purposes, demonstrating the advantages of GNN-based methodologies with respect to the state-of-the-art when including information about the underlying grid.
- Presenting an in-depth comparison of the proposed HEMS using a realistic benchmark LV distribution grid in Germany, considering a full year of electrical load and solar generation data, as well as the grid.
- Considering the uncertainty stemming from erroneous forecasts in the HEMS control problem, demonstrating the effect of practical limitations on these systems.
- Demonstrating and discussing the advantages and disadvantages of the proposed model-free and model-based methods for the first time, comparing the multiple control algorithms based on their overall performance, such as objective function values, grid constraint violations, and calculation times, and highlighting their individual strengths and weaknesses for BESS scheduling in the context of HEMS applications for DSOs.

The rest of the paper is organized as follows: Section 2 introduces the rule-based, model-based, and model-free methods used to control the BESS in practical HEMS from a DSO perspective, as well as the grid and smart building modeling. Section 3 presents the setup on which the methods are benchmarked and the results of applying the proposed control approaches, providing a detailed comparison based on the various relevant metrics. Finally, Section 4 summarizes the main findings of the paper and provides an outlook on future research work.

2. Methodology

2.1. Notation

In this paper, matrices are denoted as bold capital letters such as $\mathbf{A}$, and vectors as bold lowercase letters such as $\mathbf{a}$. Indices are denoted as non-bold, italic lowercase letters like i. Calligraphic letters such as $\mathcal{N}$ define sets. Approximated quantities are represented using the hat notation like $\hat{a}$. To denote maximum and minimum values, non-italic superscripts

are used, as for example, $E^{\max}_{\text{bat,k}}$, which represents the maximum energy of battery k.

2.2. Practical HEMS control framework

It is assumed here that building HEMS are centrally controlled in practice through their grid-connected BESS, considering the interests of both the DSO and customer, as shown in Fig. 1, which illustrates the interactions between practical building BESS and the LV grid. It consists of a Control Policy (CP), a forecaster, and an Environment, which contains the simulation models of the smart buildings and the grid. The CP represents the different control methods, which can be realized using rule-based, model-free, or model-based approaches. The respective policy acts based on the measurements recorded at time step t, which in this case consist of the State Of Charge (SOC) of all batteries in the grid $\mathbf{E}_{\text{bat}}(t)$, as well as the electrical loads $\mathbf{P}_{\text{load}}(t)$ and solar power $\mathbf{P}_{\text{pv}}(t)$ at time step t. In case of a predictive CP, the implemented forecaster provides profiles for electrical load and solar generation for a given horizon H as additional input for the CP. The CP exploits the received input data to determine suitable BESS power set points $\mathbf{P}_{\text{bat}}(t)$ as control actions for time step t for all batteries in the grid. It should be noted that only the active power of the batteries is assumed to be controllable, as the reactive power is usually not controlled by the DSO, as is the case in, for example, §14 a EnWG in Germany [40].

In the Environment in Fig. 1, the updated SOC of each battery is calculated based on the physical BESS model, and the current control actions $\mathbf{P}_{\text{bat}}(t)$. Furthermore, violations of the grid constraints are evaluated by the proposed control framework for each CP using the exchanged power $\mathbf{P}_k(t)$ with the grid as shown in Fig. 1. The aim is to compare model-based and model-free methods with respect to their performance by evaluating several metrics. For these comparisons, baseline cases are defined based on a decentralized control approach in which each building typically would have a rule-based shortsighted controller to minimize its power exchange with the LV grid, and the case where there are no BESS to control.

2.3. Environment

The Environment is the LV grid with its connected buildings, in which battery setpoints impact power flows for each discrete time step. Hence, the Environment consists of two components, namely, the grid model and the smart building model, as explained next.

2.3.1. Grid model

Electrical grids can be modeled as graphical models, where the graph $\mathcal{G}$ consists of nodes k that are part of the node set $\mathcal{N}$ and edge tuples $(k,l) \in \mathcal{N}$ that are part of the edge set $\mathcal{E}$. Each node is characterized by its connected buildings, which contain PV generation, BESS, and electrical loads. The node indices of the respective elements are collected in the sets $\mathcal{P}$, $\mathcal{B}$, and $\mathcal{L}$, respectively. The edge elements represent transformers, overhead lines, cables and switches. The nonlinear power flow equations can then be used to describe the active and reactive powers flowing through the grid edges as follows:

$$P_{ft} = g_{ff}V_f^2 + V_fV_t\left[g_{ft}\cos(\theta_f-\theta_t) + b_{ft}\sin(\theta_f-\theta_t)\right] \tag{1}$$

$$Q_{ft} = -b_{ff}V_f^2 + V_fV_t\left[g_{ft}\sin(\theta_f-\theta_t) - b_{ft}\cos(\theta_f-\theta_t)\right] \tag{2}$$

$$P_{tf} = g_{tt}V_t^2 + V_fV_t\left[g_{tf}\cos(\theta_t-\theta_f) + b_{tf}\sin(\theta_t-\theta_f)\right] \tag{3}$$

$$Q_{tf} = -b_{tt}V_t^2 + V_fV_t\left[g_{tf}\sin(\theta_t-\theta_f) - b_{tf}\cos(\theta_t-\theta_f)\right] \tag{4}$$

where V_f and V_t are the voltage magnitude at the from and to nodes and θ_f, θ_t are the respective voltage angles. The g and b values are fixed parameters of the conductances and susceptances of edge elements. Given that the powers at all nodes are known, the Newton Raphson (NR) method can be used to solve the resulting system of equations, which determines the powers flowing over each edge element, i.e., P_{ft}, P_{tf}, Q_{ft}, and Q_{tf}, as well as the complex nodal voltages. The power balance at the Point of Common Coupling (PCC) for each building can then be modeled as follows:

$$P_k + \sum_{f,t\in\mathcal{A}(k)} P_{ft} + \sum_{f,t\in\mathcal{I}(k)} P_{tf} = 0 \quad \forall k \in \mathcal{N} \tag{5}$$

$$Q_k + \sum_{f,t\in\mathcal{A}(k)} Q_{ft} + \sum_{f,t\in\mathcal{I}(k)} Q_{tf} = 0 \quad \forall k \in \mathcal{N} \tag{6}$$

where $\mathcal{A}(k)$ is the set of edge tuples where k is the from node and $\mathcal{I}(k)$ is the set of edge tuples where k is the to node.

2.3.2. Practical smart building model

The power injection P_k is the power fed into or drawn from node k and is the sum of all powers in the building:

$$P_k + P_{\text{pv},k} - P_{\text{load},k} - P_{\text{bat},k} = 0 \quad \forall k \in \mathcal{N} \tag{7}$$

where $P_{\text{pv},k}$, $P_{\text{load},k}$ and $P_{\text{bat},k}$ are the PV generation, the electrical load and battery power of the smart building connected at node k.

In the typical LV grid under consideration Q_k can be assumed to be negligible, as in the case of Germany, where a minimum power factor of 0.9 to 0.95 is mandated depending on the installed power and type of device [41]. Consequently, reactive power contributions at the household level are not considered.

The HEMS building model considered here is an off-the-shelf system currently available in the marketplace as previously discussed, in which only a BESS is controlled to supply a given uncontrolled demand, while considering PV power injections. In this context, the active battery power for the BESS of the building $P_{\text{bat},k}$ is modeled as follows:

$$E_{\text{bat},k}(t+1) = E_{\text{bat},k}(t) + \begin{cases} \frac{1}{\eta_{\text{dch}}} P_{\text{bat},k}(t)\Delta t, & P_{\text{bat},k}(t) < 0 \\ \eta_{\text{ch}} P_{\text{bat},k}(t)\Delta t, & P_{\text{bat},k}(t) \geq 0 \end{cases} \tag{8}$$

where the change in energy between $E_{\text{bat},k}(t)$ and the next discrete time step $E_{\text{bat},k}(t+1)$ depends linearly on the battery power $P_{\text{bat},k}$, as well as the charging and discharging efficiencies η_{ch} and η_{dch}, and the time interval between discrete time steps Δt. Note that $E_{\text{bat},k}$ and $P_{\text{bat},k}$ must remain within the technical limits, $E^{\min}_{\text{bat},k}$ and $E^{\max}_{\text{bat},k}$, as well as $P^{\min}_{\text{bat},k}$ and $P^{\max}_{\text{bat},k}$ at all time steps.

2.4. Control policies

As discussed in Section 1, the methods investigated for controlling the individual BESS in the considered benchmark grid are categorized into rule-based, model-based, and model-free approaches as depicted in Fig. 2.

The underlying modeling for the considered BESS CP is explained in detail next:

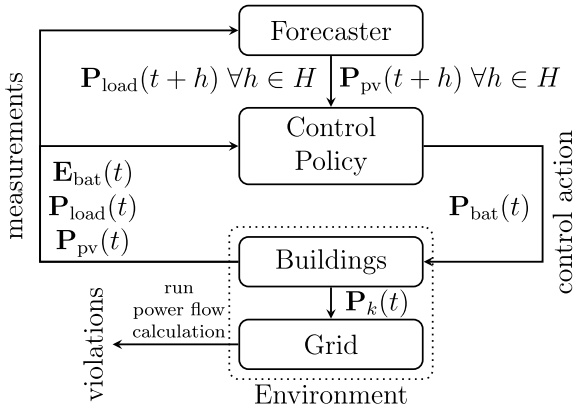


Fig. 1. Control framework for multiple grid-connected smart buildings.

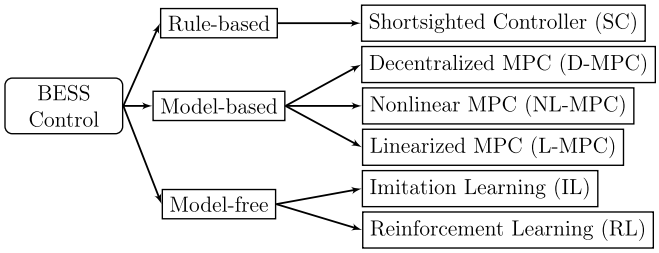


Fig. 2. Overview of applied control policies in the BESS framework.

2.4.1. Shortsighted controller (SC)

The SC is used as the baseline for comparison purposes, since it is a simple and effective state-of-the-art heuristic controller widely used in practical HEMS for the following reasons [42]:

- It is a basic controller that is grounded on decision rules that can be represented by simple equations, resulting in a straightforward implementation with low computational needs.
- It is typically based on measurements of SOC, load power, and PV power as inputs without the use of forecasting techniques.

The SC takes the measurements of SOC, load, and PV power at the current time step t into account to minimize the grid power exchange between the building at node k with the considered LV grid. The rules for determining the set points for the BESS can be defined as follows:

$$\hat{P}_{\text{bat},k}(t) = P_{\text{pv},k}(t) - P_{\text{load},k}(t) \tag{9}$$

$$P_{\text{up},k}(t) = \min\left\{\frac{E^{\max}_{\text{bat},k} - E_{\text{bat},k}(t)}{\Delta t}, P^{\max}_{\text{bat},k}\right\} \tag{10}$$

$$P_{\text{low},k}(t) = \max\left\{\frac{E^{\min}_{\text{bat},k} - E_{\text{bat},k}(t)}{\Delta t}, P^{\min}_{\text{bat},k}\right\} \tag{11}$$

$$P_{\text{bat},k}(t) = \text{Hysteresis}\left\{\hat{P}_{\text{bat},k}(t), P_{\text{low},k}(t), P_{\text{up},k}(t)\right\} \tag{12}$$

where $\hat{P}_{\text{bat},k}$ is the BESS set point that would reduce the nodal injection to zero at the PCC. To take the technical limitations of the respective battery into account, the upper and lower bounds $P_{\text{up},k}$ and $P_{\text{low},k}$ are represented in (10) and (11). The control action $P_{\text{bat},k}$ is then calculated using a hysteresis function that clips $\hat{P}_{\text{bat},k}$ within these bounds. Note that the SC does not take forecasts of future power consumption and generation nor information on grid variables, i.e., voltages and currents, into account. While it promptly adjusts to system changes, its lack of optimal oversight may lead to inefficiencies [7].

2.4.2. Model predictive control (MPC)

Battery schedules are optimized here for peak shaving, which is a primary objective for DSOs and also decreases the homeowners electricity costs. In this context, a rolling-horizon MPC is formulated to solve the optimization problem with fixed system constraints at each time step t based on measurements, PV generation, and load demand forecasts for a given horizon length H, with a temporal resolution Δt. Only the set point for the first time step from each optimal battery schedule is used, thus mitigating forecasting errors. This procedure is repeated for each time step, taking future changes in the system into account [7]. Two different centralized MPC formulations are proposed, i.e., a Nonlinear MPC (NL-MPC) formulation considers all buildings and nonlinear formulations of the LV grid equations in the optimization problem. On the other hand, a Linearized MPC (L-MPC) formulation is based on linearized versions of the nonlinear grid equations. Additionally, a Decentralized MPC (D-MPC) is implemented using a separate HEMS MPC formulation for each building without considering the grid. All proposed MPCs are introduced in detail next.

The NL-MPC takes nonlinear grid equations into account and solves the following centralized optimization problem at each time step t:

$$\min_{P_{\text{bat},k}} \quad \sum_{t=1}^{H}\sum_{k\in\mathcal{N}} P_k(t)^2 \tag{13}$$

$$\text{s.t.} \quad \forall k \in \mathcal{N},\ \forall (k,l) \in \mathcal{E},\ \forall t \in H$$

$$P_{kl}(t) = g_{kk}V_k^2(t) + V_k(t)V_l(t)$$
$$\left[g_{kl}\cos(\theta_k(t) - \theta_l(t)) + b_{kl}\sin(\theta_k(t) - \theta_l(t))\right] \tag{14}$$

$$Q_{kl}(t) = -b_{kk}V_k^2(t) + V_k(t)V_l(t)$$
$$\left[g_{kl}\sin(\theta_k(t) - \theta_l(t)) - b_{kl}\cos(\theta_k(t) - \theta_l(t))\right] \tag{15}$$

$$P_{lk}(t) = g_{ll}V_l^2(t) + V_l(t)V_k(t)$$
$$\left[g_{lk}\cos(\theta_l(t) - \theta_k(t)) + b_{lk}\sin(\theta_l(t) - \theta_k(t))\right] \tag{16}$$

$$Q_{lk}(t) = -b_{ll}V_l^2(t) + V_l(t)V_k(t)$$
$$\left[g_{lk}\sin(\theta_l(t) - \theta_k(t)) - b_{lk}\cos(\theta_l(t) - \theta_k(t))\right] \tag{17}$$

$$\text{Re}\left\{I_{kl}(t)\right\} = g_{kk}V_k(t)\cos\left(\theta_k(t)\right) - b_{kk}V_k(t)\sin\left(\theta_k(t)\right) + g_{kl}V_l(t)\cos(\theta_l(t)) - b_{kl}V_l(t)\sin(\theta_l(t)) \tag{18}$$

$$\text{Im}\left\{I_{kl}(t)\right\} = b_{kk}V_k(t)\cos\left(\theta_k(t)\right) + g_{kk}V_k(t)\sin\left(\theta_k(t)\right) + g_{kl}V_l(t)\sin(\theta_l(t)) + b_{kl}V_l(t)\cos(\theta_l(t)) \tag{19}$$

$$\text{Re}\left\{I_{lk}(t)\right\} = g_{ll}V_l(t)\cos\left(\theta_l(t)\right) - b_{ll}V_l(t)\sin\left(\theta_l(t)\right) + g_{lk}V_k(t)\cos(\theta_k(t)) - b_{lk}V_k(t)\sin(\theta_k(t)) \tag{20}$$

$$\text{Im}\left\{I_{lk}(t)\right\} = b_{ll}V_l(t)\cos\left(\theta_l(t)\right) + g_{ll}V_l(t)\sin\left(\theta_l(t)\right) + g_{lk}V_k(t)\sin(\theta_k(t)) + b_{lk}V_k(t)\cos(\theta_k(t)) \tag{21}$$

$$P_k(t) + P_{\text{pv},k}(t) - P_{\text{load},k}(t) - P_{\text{bat},k}(t) = 0 \tag{22}$$

$$Q_k(t) + Q_{\text{pv},k}(t) - Q_{\text{load},k}(t) - Q_{\text{bat},k}(t) = 0 \tag{23}$$

$$P_k(t) + \sum_{m,n\in\mathcal{A}(k)} P_{mn}(t) + \sum_{m,n\in\mathcal{I}(k)} P_{nm}(t) = 0 \tag{24}$$

$$Q_k(t) + \sum_{m,n\in\mathcal{A}(k)} Q_{mn}(t) + \sum_{m,n\in\mathcal{I}(k)} Q_{nm}(t) = 0 \tag{25}$$

$$E_{\text{bat},k}(t+1) = E_{\text{bat},k}(t) + P_{\text{bat},k}(t)\Delta t \quad \forall k \in \mathcal{B}\ \forall t \in H \tag{26}$$

$$P^{\min}_{\text{bat},k} \le P_{\text{bat},k}(t) \le P^{\max}_{\text{bat},k} \quad \forall k \in \mathcal{B} \quad \forall t \in H \tag{27}$$

$$E^{\min}_{\text{bat},k} \le E_{\text{bat},k}(t) \le E^{\max}_{\text{bat},k} \quad \forall k \in \mathcal{B} \quad \forall t \in H \tag{28}$$

$$V^{\min} \le V_k(t) \le V^{\max} \tag{29}$$

$$\sqrt{\text{Re}\left\{I_{kl}(t)\right\}^2 + \text{Im}\left\{I_{kl}(t)\right\}^2} \le I^{\max}_{kl} \tag{30}$$

$$\sqrt{P^2_{\text{trf}}(t) + Q^2_{\text{trf}}(t)} \le S^{\max}_{\text{trf}} \tag{31}$$

where the nonlinear objective (13) leads to peak shaving, as proposed in for example [43] and [44], incentivizing the reduction of peaks in grid power injections at their respective PCC for all nodes and time

steps. This leads to both the reduction of grid congestion and an increase in self-sufficiency of the grid-connected buildings, which results in reduced electricity costs for the consumers. For each edge element, the active and reactive power flows at both terminals and for all horizon time steps are represented by the nonlinear power flow equations in (14)–(17). The real and imaginary currents for all edge elements are modeled in (18)–(21). The nodal power balance for active and reactive power is taken into account in (22)–(25). The voltage magnitudes in the grid are bound between $V^{\min}$ and $V^{\max}$ for all nodes and time steps in (29). The line current $I_{kl}(t)$ is bounded by the thermal rated current $I_{kl}^{\max}$ in (30). To ensure that the transformer connecting the smart grid with the Medium Voltage (MV) grid is not overloaded, (31) bounds the apparent power by the maximum rated power of the transformer $S_{\text{trf}}^{\max}$. The battery represents the flexible component of the system and can be freely dispatched within its technical limits, which are considered in constraints (26)–(28). For computational efficiency, the SOC of the battery at time step $t+1$ is approximated using a linear constraint, assuming that charging and discharging efficiencies are similar and relatively high, i.e., close to 100 %, as assumed here; otherwise, the model would require binary variables to represent both the BESS charging and discharging, thus resulting in a complex mixed-integer programming problem.

The grid Eqs. (14)–(21) are nonlinear and lead to an NLP optimization problem that consists of a quadratic objective function (13), as well as eight nonlinear flow constraints for each edge element in $\mathcal{E}$. Hence, an L-MPC approach is used here based on the Linear-Coupled Power Flow (LCPF) equations introduced in [45], to derive a simplified linear optimization problem that can be solved faster introducing small errors from the linearization. In this case, the optimization problem solved in each step of the L-MPC can be formulated as follows:

$$\min_{P_{\text{bat},k}} \quad \sum_{t=1}^{H} \sum_{k\in\mathcal{N}} P_k(t)^2 \tag{32}$$

$$\text{s.t.} \quad \forall k \in \mathcal{N}, \quad \forall (k,l) \in \mathcal{E}, \quad \forall t \in H$$

$$P_{kl}(t) = -P_{lk}(t) = g_{kl}\big(V_k(t) - V_l(t)\big) - b_{kl}\big(\theta_k(t) - \theta_l(t)\big) \tag{33}$$

$$Q_{kl}(t) = -Q_{lk}(t) = -b_{kl}\big(V_k(t) - V_l(t)\big) - g_{kl}\big(\theta_k(t) - \theta_l(t)\big) \tag{34}$$

$$P_k(t) + P_{\text{pv},k}(t) - P_{\text{load},k}(t) - P_{\text{bat},k}(t) = 0 \tag{35}$$

$$Q_k(t) + Q_{\text{pv},k}(t) - Q_{\text{load},k}(t) - Q_{\text{bat},k}(t) = 0 \tag{36}$$

$$P_k(t) + \sum_{m,n}^{\mathcal{A}(k)} P_{mn}(t) + \sum_{m,n}^{\mathcal{I}(k)} P_{nm}(t) = 0 \tag{37}$$

$$Q_k(t) + \sum_{m,n}^{\mathcal{A}(k)} Q_{mn}(t) + \sum_{m,n}^{\mathcal{I}(k)} Q_{nm}(t) = 0 \tag{38}$$

$$E_{\text{bat},k}(t+1) = E_{\text{bat},k}(t) + P_{\text{bat},k}(t)\Delta t \quad \forall k \in \mathcal{B}\ \forall t \in H \tag{39}$$

$$V^{\min} \le V_k(t) \le V^{\max} \tag{40}$$

$$-I_{kl}^{\max} \le \alpha P_{kl}(t) \le I_{kl}^{\max} \tag{41}$$

$$P_{\text{bat},k}^{\min} \le P_{\text{bat},k}(t) \le P_{\text{bat},k}^{\max} \quad \forall k \in \mathcal{B} \quad \forall t \in H \tag{42}$$

$$E_{\text{bat},k}^{\min} \le E_{\text{bat},k}(t) \le E_{\text{bat},k}^{\max} \quad \forall k \in \mathcal{B} \quad \forall t \in H \tag{43}$$

$$P_{\text{trf}}^{\max} \le P_{\text{trf}}(t) \le P_{\text{trf}}^{\max} \tag{44}$$

Thus, Eqs. (14)–(17) in the NL-MPC optimization problem are replaced by the linearized power flow equations in (33) and (34) which introduce small deviations in the voltages and currents. This results in linear constraints for the optimization problem, which aims at finding BESS schedules that result in peak shaving over horizon H, while satisfying all required constraints. Note that the respective powers must be converted into appropriate current values in (41) to comply with the current limits, which can be readily accomplished with the factor $\alpha = \frac{S_{\text{nom}}\cdot 10^6}{\sqrt{3}V_{\text{nom}}\cdot 10^3}$, where S_{nom} and V_{nom} are the nominal apparent power in MVA and the nominal voltage of the respective edge element in kV.

In most cases, the LCPF model converges within the given limits $V^{\max}$, $V^{\min}$, $P_{\text{trf}}^{\max}$, $I_{kl}^{\max}$; however in case of no convergence, the bounds in (40), (41), and (44) are iteratively relaxed to facilitate power flow convergence. Note that the resulting optimization problem contains linear equality and inequality constraints, while the objective function is quadratic, thus resulting in a Linear Constrained Quadratic Programming (LCQP) problem.

Finally, a D-MPC approach is also considered to allow evaluating its impact on the grid. This assumes that each HEMS solves a separate MPC, taking only the model of the BESS, household, and PCC into account. In this case, only objective (13) and constraints (7), (26)–(28) are considered for each MPC, since the grid equations are not included, effectively decoupling the optimization problems for each HEMS.

2.5. Imitation learning and graph imitation learning

IL aims at learning a CP by imitating the behavior of a model-based MPC. In this case, an MPC algorithm is first simulated for a specified number of time steps. The results of the MPC simulation are divided into input data vectors $\mathbf{X}(t) = [E_{\text{bat},k}(t), P_{\text{res},k}(t), P_{\text{res},k}(t+1), \dots, P_{\text{res},k}(t+H)\ \forall k \in \mathcal{N}]$, with $P_{\text{res},k}(t) = P_{\text{pv},k} - P_{\text{load},k}(t)$, and output data vectors $\mathbf{Y}(t) = [P_{\text{bat},k}(t)\ \forall k \in \mathcal{B}]$, which are concatenated for all simulated time steps T, resulting in the matrices $[\mathbf{X}(1), \dots, \mathbf{X}(T)]$ and $[\mathbf{Y}(1), \dots, \mathbf{Y}(T)]$. The aim is then to train an arbitrary function approximator f_θ that learns a parameterizable CP for the following functional mapping:

$$\hat{\mathbf{Y}}^{\text{IL}}(t) \leftarrow f_\theta^{\text{IL}}(\mathbf{X}(t)) \quad \forall t \in T \tag{45}$$

where $\hat{\mathbf{Y}}^{\text{IL}}(t)$ is the vector of approximated control actions at time step t, and θ represents the weights and biases of the chosen NN. Depending on the selected regression method, the approximator can then be trained on the control outputs of the model-based MPC method with the objective of reducing the following Mean Squared Error (MSE) loss between the control outputs of both methods:

$$\frac{1}{n_{\text{samples}}} \sum_{t=1}^{T} \sum_{k\in\mathcal{B}} \big(\hat{Y}_k^{\text{IL}}(t) - Y_k^{\text{MPC}}(t)\big)^2 \tag{46}$$

where $n_{\text{samples}} = T|\mathcal{B}|$ is the total number of control actions chosen by the MPC. The IL setup in (45) does not contain information about the underlying physical system, since only the mapping from measurements of SOC, PV, and load, as well as forecasts of PV and load are taken into account to learn a mapping to the respective control actions.

The information of the underlying grid in form of a graph $\mathcal{G}$ can be included in the regression problem using Geometric Imitation Learning (GIL), leading to:

$$\hat{\mathbf{Y}}^{\text{GIL}}(t) \leftarrow f_\theta^{\text{GIL}}(\mathbf{X}(t), \mathcal{G}) \quad \forall t \in T \tag{47}$$

where $\hat{\mathbf{Y}}^{\text{GIL}}(t)$ is the vector of approximated control actions at time step t and f_θ^{GIL} is the respective policy. Note that this GIL regression problem requires specialized regression methods that are able to take the graph $\mathcal{G}$ into account, such as Geometric Deep Learning (GDL) methods.

Fig. 3 shows the proposed IL framework that is applied in this paper, where NNs are used as function approximators due to their scalability properties and the possibility of considering the graph in the regression, and $f_{\mathcal{G}}^{\text{MPC}}$ stands for the NL-MPC or L-MPC centralized approaches, which account for the grid $\mathcal{G}$.

The loss in (46) is used for training, where the weights and biases of the NNs are updated using stochastic gradient descent. The layer type GraphSAGE in [46] of the GNN is chosen for the GIL implementation, since this is suitable for GDL tasks in power systems as shown in [47]. For IL, NNs based on Multi-Layer Perceptrons (MLP) are used.

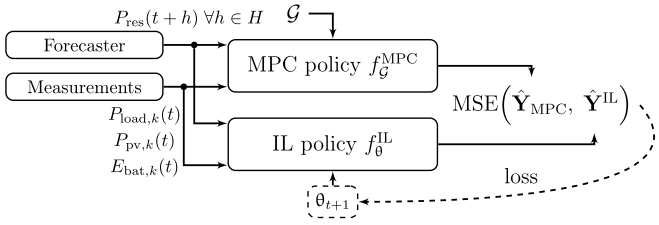


Fig. 3. Imitation learning framework for IL and GIL.

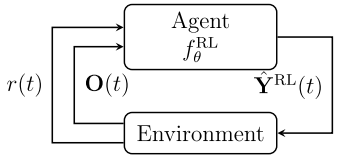


Fig. 4. Reinforcement learning framework for BESS scheduling.

2.6. Reinforcement learning

RL aims at learning a CP f^{RL} through repeated interactions with an Environment, as shown in Fig. 4, where the generic RL interaction scheme between simulation Environment and agent is depicted. After each action $\hat{\mathbf{Y}}^{RL}(t)$ chosen by the RL agent, the Environment returns a new observation $\mathbf{O}(t)$ and the corresponding reward $r(t)$, which should be maximized by the agent over time. A black-box function approximator, such as NNs, is usually applied to represent the control policy of the agent. Unlike IL, the RL approach is not a supervised learning task and no ground truth generated by an MPC is required, resulting in flexible algorithms that do not require system modeling. To enable the application of RL to control the BESS, an interaction scheme is used between an agent and the Environment, which in this case is a simulation model of the HEMS and the LV grid. Thus, the agent f_{θ}^{RL} interacts with the Environment to maximize the following scalar reward $r(t)$:

$$r(t) = -\lambda_{c_1} c_1(t) - \lambda_{c_2} c_2(t) - \lambda_{c_3} c_3(t) - \lambda_{c_4} c_4(t) - \lambda_{c_5} c_5(t) \tag{48}$$

$$c_1(t) = \sum_{k \in \mathcal{N}} \left(P_{\text{load},k}(t) + P_{\text{bat},k}(t) - P_{\text{pv},k}(t)\right)^2 \tag{49}$$

$$c_2(t) = \sum_{k \in \mathcal{N}} \max(V_k(t) - V^{\max}, 0) + \max(V^{\min} - V_k(t), 0) \tag{50}$$

$$c_3(t) = \sum_{k,l \in \mathcal{E}} \max\left(\left(\max\left(|I_{kl}(t)|, |I_{lk}(t)|\right) - I_{kl}^{\max}\right), 0\right)^2 \tag{51}$$

$$c_4(t) = \left(\max(S_{\text{trf}}(t) - S_{\text{trf}}^{\max}, 0) + \max(-S_{\text{trf}}^{\max} - S_{\text{trf}}(t), 0)\right)^2 \tag{52}$$

$$c_5(t) = c_t \tag{53}$$

where (48) is a sum of five reward components c_1, c_2, c_3, c_4 and c_5, which can be weighted using the fixed factors λ_{c_1} to λ_{c_5}. Reward component c_1 in (49) penalizes large outliers, similar to the peak shaving objective in (13). Voltages that exceed the maximum allowed voltage $V^{\max}$, as well as voltages that are lower than the minimum allowed voltage $V^{\min}$ are penalized using (50). Penalties for exceeding maximum current rating of lines and power rating of the transformer are represented in (51) and (52), respectively. In cases where a non-convergent load flow is caused by the actions chosen by the agent, component c_5 in (53) is active, where c_t is a binary variable equal to 1 when the power flow calculation in the simulation environment does not converge. Note that $r(t)$ contains the objective used by the MPC, as well as penalty terms for constraint violations, i.e., voltage bounds and line constraints, which can only be taken into account implicitly.

Table 2
Considered model-free RL methods.

Algorithm	Source
Augmented Random Search (ARS)	[48]
Deterministic Policy Gradients (DDPG)	[49]
Proximal Policy Optimization (PPO)	[50]
Soft Actor Critic (SAC)	[51]
Twin-Deterministic DDPG (TD3)	[52]
Asynchronous Advantage Actor Critic (A2C)	[53]

The goal of the agent is to find a control policy f^{RL} that maximizes the discounted total reward $\tilde{r}$:

$$\tilde{r} = \sum_{t=1}^{\infty} \gamma_t r(t) \tag{54}$$

where $0 < \gamma_t < 1$ is a discount factor that ensures that the sum in (54) is finite, with reward values that are further in the past having less weight in the calculation of $\tilde{r}$. Similarly to IL in (45), the policy is a functional mapping from observations to control actions:

$$\hat{\mathbf{Y}}(t)^{\text{RL}} \leftarrow f_{\theta}^{\text{RL}}(\mathbf{O}(t)) \tag{55}$$

where f_{θ}^{RL} is approximated by an NN, which obtains the observations $\mathbf{O}(t)$ at time step t and calculates the corresponding control action vector $\hat{\mathbf{Y}}(t)^{\text{RL}}$ based on these inputs. The observations are a vector of measured and forecasted values at time step t and defined as follows for the given control task:

$$\tilde{\mathbf{O}}(t) = \left[\ \mathbf{E}_{\text{bat}}(t),\ \mathbf{P}_{\text{res}}(t)\ \right] \tag{56}$$

$$\mathbf{O}(t) = \left[\ \mathbf{E}_{\text{bat}}(t),\ \mathbf{P}_{\text{res}}(t), \mathbf{P}_{\text{res}}(t+1), \ldots, \mathbf{P}_{\text{res}}(t+H)\ \right] \tag{57}$$

where $\tilde{\mathbf{O}}(t)$ is the reduced observation at time step t, consisting of the measured SOC of all batteries in $\mathbf{E}_{\text{bat}}(t)$ and the residual power at all nodes in $\mathbf{P}_{\text{res}}(t)$. Similarly, $\mathbf{O}(t)$ is the full observation, containing $\tilde{\mathbf{O}}(t)$, as well as forecasts for the residual powers for the next H time steps. The policy is then iteratively updated based on the considered state-of-the-art RL algorithms shown in Table 2. The discounted total reward in (54), which the agent receives when interacting with the environment, has to be maximized by the agent over time.

2.7. Forecaster

The forecaster is part of the framework in Fig. 1 and obtains the measurements of the PV and load power at each time step t. Based on these inputs, a forecast of PV and load power for the horizon H with resolution Δt is calculated by the forecaster. Forecasting of power system time-series is a broad topic that is covered extensively in the scientific literature (e.g., [54]), applying methods from statistics and Machine Learning (ML). Since this work focuses on investigation, different, simplified assumptions are made regarding the forecaster to cover the possible range of forecast accuracies that might arise in real-world applications. Therefore, two forecasting algorithms are used here:

- A Worst-Case Forecaster (WCF) that is based on the naive assumption that the profiles from the last 24 h are repeated for the next day, thus representing the forecasting algorithm often used as baseline in the literature [55].
- A Best-Case Forecaster (BCF) that assumes perfect forecasts, resulting in a lower bound on forecasting errors for PV and load powers.

By applying WCF and BCF in the framework in Fig. 1, the range of possible attainable results can be covered, assuming that in reality the forecaster will be between both forecasting approaches. Note that both WCF and BCF provide deterministic point-forecasts, which is the state-of-the-art for HEMS [14]. Furthermore, probabilistic approaches come with the disadvantage of high calculation times, which lead to extensive simulation times for the study spanning one year conducted in this paper since, 34,848 simulations are required for each simulated year.

3. Results

3.1. Simulation setup

In this paper, a benchmark LV grid is considered with 13 different, realistic smart buildings connected, incorporating electric household loads, PV generation, and controllable BESS. The applied load profiles are the same as in [11], which are based on realistic simulations of residential behavior from [56] with a time resolution of $\Delta t = 15$ min. The PV profiles are derived from actual feed-in data collected from PV plants located in Karlsruhe, Germany. These profiles are adjusted to align with the annual energy consumption of the corresponding household loads. Each BESS is sized according to the methodology outlined in [57] to enhance flexibility within the system, assuming a C-rate of one. The PV generation and demand forecasts for the next 24 hours every 15 min result in a horizon $H = 96$, with forecasts based on the BCF or WCF approaches.

The German benchmark LV grid "1-LV-rural1-1-sw" is used here, based on the `SimBench` Python package [58], which provides well-established benchmark grids for Germany. This grid comprises 15 nodes and 13 lines with smart buildings connected at a 400 V level and is depicted in Fig. 5. The transformer linking the 400 V LV grid with the 20 kV MV grid is rated at 0.16 MVA. To assess the effects of the examined CPs, the load and PV profiles were scaled up to force grid limit violations without the usage of BESS.

The software packages used for the proposed framework were `Pandapower` for power flows [59], and `Pyomo` with the IPOPT solver for optimization [60]. IL and GIL were implemented using `PyTorch` [61] and `PyTorchGeometric` [62], respectively. For the RL algorithms, `Stable Baselines3` [63] was used for the developed environments. All algorithms and simulations were run on a computer with a Threadripper 3990X 64×2.9 GHz CPU, 128 GB RAM, and a Nvidia TITAN RTX GPU with 4608 CUDA cores.

The developed evaluation procedure used for the proposed CPs starts with grid, technical constraints, load, and PV data being provided for the testing period, as illustrated in Fig. 6. At each time step, the measurements and forecasts are simulated for the corresponding CP. Depending on the selected CP, new BESS set points for each smart building are determined. Finally, the BESS set points are used to simulate the changes in SOC with the battery model in (8), denoted with the developed `physical` function, and to calculate the required power flows. The BESS efficiencies η_{ch} and η_{dch} are assumed based on commercially available data for home storage systems with lithium-ion batteries [64,65]. Based on the results of the grid calculations, the specified metrics for the respective CP can be determined. This procedure is simulated for the

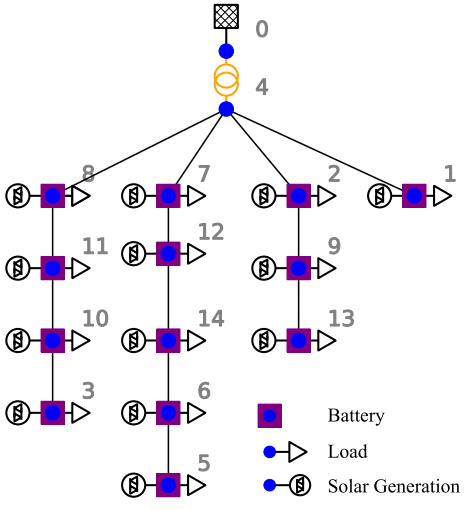


Fig. 5. German benchmark LV grid.

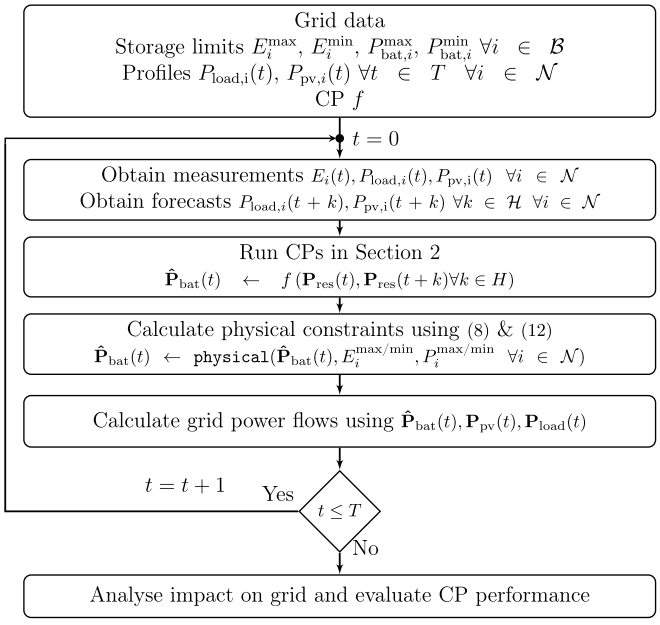


Fig. 6. CP evaluation procedure.

complete testing dataset, which consists of one year with $T = 34,848$ time steps. Note that, since no actual data is available to test the proposed ML techniques, given the novelty of the proposed approaches, the presented results are based on the synthetic data obtained in the evaluation procedure illustrated in Fig. 6, as is typically done in these cases (e.g., [66,67]).

3.2. Evaluation metrics

The chosen metrics aim at quantifying the performance of each CP from the perspective of both the homeowners and DSO. The peak shaving ability is quantified using the following value, as proposed in [43,68]:

$$J_{\text{ps}} = \sum_{t=1}^{T} \sum_{k \in \mathcal{N}} P_k(t)^2 \tag{58}$$

which is similar to (13), but calculated for all simulated time steps T of the testing dataset. The number of constraint violations was counted, assuming $V^{\min} = 0.9$ p.u. and $V^{\max} = 1.1$ p.u. for all nodes and at each time step, as per German LV standards, $I_{kl}^{\max} = 0.142$ kA for the LV grid used, and $S_{\text{trf}}^{\max} = 0.16$ MVA for the MV/LV transformer. The calculation time required to execute the policy during operation was tracked using the mean value μ and standard deviation σ to evaluate the potential practical application of the CP.

3.3. Model predictive control evaluation

For the evaluation of the rolling-horizon approach, the MPC optimization problem is solved after each time step Δt with updated forecasts and new measurements as input data. Hence, over the course of a one-year rolling-horizon MPC, a total of 34,848 optimization problems were solved. In case of an infeasible optimization problem, an iterative relaxation scheme for the grid constraints was implemented, which was applied until a feasible solution was found. Three MPC methods were simulated with the testing dataset spanning one year for comparison purposes, i.e., NL-MPC based on (13)–(31), L-MPC based on the linearized formulation (33)–(44), and D-MPC with BESS constraints. Both BCF and WCF were considered in separate simulations, leading to six different configurations.

Table 3
Hyperparameters of IL and GIL neural networks.

Model	NN Architecture	Hyperparameters
IL	n=200 n=200 n=200 x → Dense → ReLU → Dense → ReLU → Dense → y $\mathbf{x} \in \mathbb{R}^{(H+1)\cdot\lvert\mathcal{B}\rvert\times 1}$ $\mathbf{y} \in \mathbb{R}^{\lvert\mathcal{B}\rvert\times 1}$	• batch size 8 • epochs 60 • learning rate 5×10^{-4} • SGD optimizer
GIL	n=970 n=970 n=970 x → SAGE → ReLU → SAGE → ReLU → SAGE → y $\mathcal{G}$ $\mathcal{G}$ $\mathcal{G}$ $\mathbf{x} \in \mathbb{R}^{\lvert\mathcal{B}\rvert\times(H+1)}$ $\mathbf{y} \in \mathbb{R}^{\lvert\mathcal{B}\rvert\times 1}$	• 10 epochs • 97 node input feats • learning rate 5×10^{-5} • weight decay 1×10^{-7} • dropout 0.1 • Adam optimizer

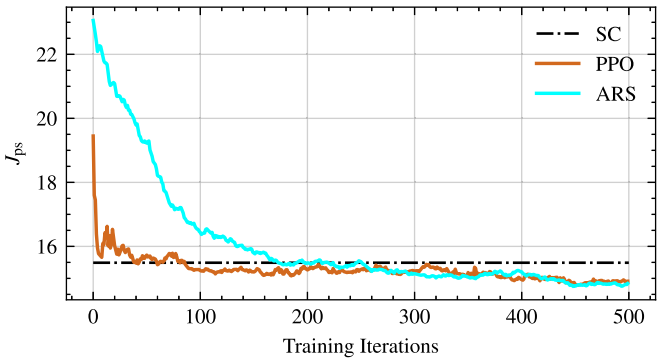


Fig. 7. Objective function values J_{ps} during RL training on 500 iterations with PPO and ARS.

3.4. Imitation learning evaluation

Two control settings are compared, i.e., IL using a NN with MLP-architecture, and GIL using a NN with GraphSAGE layers from the `PyTorchGeometric` package. The GIL approach considers an undirected graph $\mathcal{G}$ of the grid, consisting of the topology, as well as the individual edge conductances g_{kl} and susceptances b_{kl}. Table 3 shows the best hyperparameters that were found through line search for both approaches. For training IL and GIL, a separate training dataset is used. It consists of the previous year of data, based on the same buildings, BESS and PV generators. The training dataset is generated by the NL-MPC with BCF, and training IL and GIL takes 8 and 38 minutes, respectively. The testing dataset comprises the subsequent, unseen year of data and the trained NNs of f^{IL} and f^{GIL} are used to control the BESS in the evaluation stage, using both BCF and WCF forecasts as input.

3.5. Reinforcement learning evaluation

RL results were obtained using the framework in Fig. 1. Nonlinear policies such as different NN architectures as well as linear policies were considered for f^{RL}. All algorithms in Table 2 were trained until convergence was observed, which took around 500 training iterations using the same training data. Note that these iterations correspond to 500 repetitions of one year of training data, which is not practical in real-time, as this would require extensive simulations of the proposed system controls to generate the required years of data, or alternatively, measurements over multiple years of the proposed controller. Each training iteration consists of multiple steps, including simulating the environment, calling the RL controller, and updating the weights in the RL controller, which is the most time-consuming part of the process. Once the RL training is completed, the controller quickly provides set points without the need for retraining.

The state-of-the-art, model-free RL methods shown in Table 2 were compared using different implementations of the reward, the observation space, and various hyperparameters. While some improvement was observed over the course of the training, all model-free RL methods did not find policies that were competitive with the other approaches, due to the high number of parameters in the NN and the complexity of the BESS scheduling task in the HEMS. Therefore, modifications were made to obtain the most competitive RL policies, which surpassed the results of SC, resulting in the following modified RL methods:

- ARS with a linear policy f^{RL} based on the reduced observation space $\tilde{\mathbf{O}}(t)$, which only takes values at time step t into account.
- PPO with a nonlinear policy f^{RL} based on $\tilde{\mathbf{O}}(t)$ using a NN with the augmented features, which capture information about the future in a compressed way such as the squared residual power, the sum over all forecasted residual powers and the sum over all squared forecasted residual powers as additional input data.

The training procedure of the best run of the ARS and PPO methods shown in Fig. 7 was evaluated with respect to the values of J_{ps}. Each iteration corresponds to a full year of data the RL agent is simulated on, which took 100 min for the ARS and about 13.5 h for the PPO. As a reference, the result of the SC is depicted as a horizontal line. Both methods exceed the J_{ps} values obtained with SC after 500 iterations of training data. The results of the other RL methods are highly heterogeneous depending on the method and the simulation run. It is important to note that NNs with larger architectures consisting of multiple hidden layers and neurons, resulting in a higher number of trainable parameters, i.e., weights and biases, were tested with smaller NNs performing better by showing greater improvements over the training period. Hence, given the time constraints and the poor convergence of larger NNs, policies utilizing smaller NNs were used.

RL methods are non-deterministic methods that can get stuck in local minima or whose performance can deteriorate over time, as in the case of A2C, which experiences a strong deterioration in relation to the reward obtained in the simulation run considered. Note also that, while the RL framework limits the power set points of the BESS, the technical limitations of the grid are not taken into account as constraints, which might lead to unsafe actions that can violate grid limits. Overall, it can be concluded that although RL can learn methods that are better than SC for J_{ps}, the training of the methods is very sensitive to the choice of hyperparameters, the definition of the reward function, and the observation space, as well as to the particular simulation run due to the non-deterministic behavior of RL. For practical applications, this means that special attention must be paid to infeasible and suboptimal control decisions of RL controllers, which may result in limit violations and suboptimal costs. Therefore, an additional check is needed to discard infeasible control set points exceeding the technical limits of the BESS, forcing the system to operate at its closest limit in this case.

3.6. Comparison of control policies

To compare the developed BESS control policies f, each policy was evaluated on a previously unseen year of data, and the results were analyzed with respect to the metrics described in Section 3.2. For the evaluation of model-free approaches, the inference time of the policy was used to obtain realistic computational times for practical applications.

The resulting metrics are given in Table 4, highlighting in bold the best results for each metric. Observe that all methods improve J_{ps} compared to the reference case without batteries. The best results are achieved for the BCF by D-MPC, L-MPC, and NL-MPC, demonstrating an improvement of 51.2 % for BCF and 21.7 % for WCF compared to the SC. This is due to the combination of perfect predictions and model-based methods. Furthermore, the heuristic, rule-based SC performs worse than the model-based and model-free policies, as no predictions are considered in this case and the correlations in the given data cannot be learned

Table 4
Overview of result metrics for the testing year of BESS scheduling.

Category	Control policy	Forecaster	$\sum J_{ps}$	voltages	transformer	lines	Calculation time per step	
			[MWh^2]	[# of violations]	[# of violations]	[# of violations]	μ [s]	σ [s]
Rule-based	No batteries	–	23.15	10	1553	16	–	–
	SC	–	15.48	10	1106	10	$\mathbf{6.0 \cdot 10^{-5}}$	$\mathbf{\pm 2.9 \cdot 10^{-5}}$
Model-based	D-MPC	WCF	12.10	4	375	6	0.71	±0.06
		BCF	**7.55**	**0**	11	**0**	0.69	±0.06
	L-MPC	WCF	12.11	4	292	6	2.42	±4.00
		BCF	**7.55**	**0**	9	**0**	2.20	±0.21
	NL-MPC	WCF	12.11	2	273	4	5.12	±9.70
		BCF	**7.55**	**0**	**0**	**0**	4.66	±0.24
Model-free	IL	WCF	12.04	6	420	5	$8.86 \cdot 10^{-4}$	$\pm 3.8 \cdot 10^{-4}$
		BCF	8.60	2	36	**0**		
	GIL	WCF	12.05	4	412	7	$3.1 \cdot 10^{-3}$	$\pm 2.3 \cdot 10^{-3}$
		BCF	7.96	**0**	12	**0**		
	RL - lin ARS	–	14.54	10	902	10	$2.1 \cdot 10^{-4}$	$\pm 9.8 \cdot 10^{-4}$
	RL - PPO	WCF	15.55	10	1097	5	$14.23 \cdot 10^{-2}$	$\pm 8.84 \cdot 10^{-4}$
		BCF	14.80	8	1093	2		

in the model. The objective function value J_{ps} of the model-free methods results in an improvement of 6.1 % for RL, 44.4 % for IL, and 48.6 % for GIL compared to the SC. This shows that RL only achieves slight improvements, while IL and GIL come close to the results of the model-based methods. This can be explained by the fact that RL has difficulties in learning the complex relationships with a continuous action space despite many training epochs, while IL and GIL are trained on the optimal schedules of the model-based methods. It is important to note that both IL and GIL are inherently limited in their control performance by the quality of the underlying CP that is being imitated, and thus the choice of the expert policy used is important.

Note in Table 4 that the amount of grid violations for all control policies is less than or equal to the reference case, which shows that the BESS results in an improvement in grid loading under the proposed peak shaving objective. The total number of grid violations is highest for the rule-based SC, as it has no knowledge of the grid equations and no predictions. As expected, the "ideal" NL-MPC for BCF, which does not result in any grid violations due to the consideration of the nonlinear load flow equations, performs best for all the considered metrics. Observe also that the consideration of linearized load flow equations for L-MPC leads to fewer overloads and grid violations than for D-MPC, showing that considering grid equation constraints is beneficial. The model-free methods achieve an improvement from 18.1 % to 99 % for the total number of grid violations compared to the rule-based SC, and are close to the results of the model-based methods.

The results in Table 4 further illustrate the influence of forecast quality on the objective function values for all policies, showing that the achievable values and grid violations are highly dependent on the accuracy of the forecasts. Notably, forecast errors in WCF significantly impact model-based policies more than model-free approaches, primarily due to the constraints inherent in the model-based MPC strategies.

The mean values μ and standard deviations σ of the required calculation time per time step in Table 4 show that the calculation time is the same for the model-free approaches with WCF and BCF, since the same trained model is used in both cases. Observe that the SC is the fastest, as it only has to calculate a hysteresis function (12). As expected, the model-based methods have the highest computation time, with the NL-MPC taking the longest due to the significant complexity of the nonlinear constraints in the optimization problem. As expected, a significant reduction in computing time can be achieved by using model-free policies, resulting in speed ups of up to four orders of magnitude over the model-based methods and only one order of magnitude slower than the rule-based SC.

Fig. 8 shows the grid power $P_{MV/LV}$ at the MV/LV transformer for a summer day with the highest irradiance for all implemented policies,

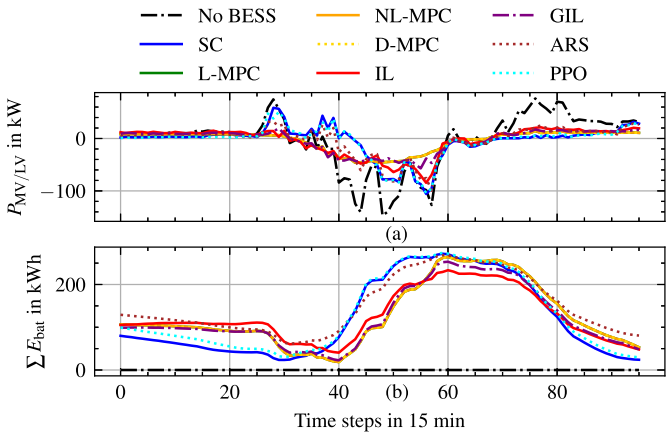


Fig. 8. Exchanged (a) power and (b) energy with MV grid for a summer day based on BCF.

with the positive and negative values indicating the flow direction of the power flow through the transformer, i.e., positive values represent power flows from the MV to the LV grid, and the opposite for negative values. The summed energy $\sum_{k \in \mathcal{B}} E_{bat,k}(t)$ of all BESS in the considered grid is also depicted. Observe that at noon, power is generated from the PV, resulting in the batteries being charged simultaneously, and the surplus yields a reverse power flow. Note also the dominant peaks in the grid power for the reference case, which are partially prevented by the SC, but not for all peaks, as the BESS are at their SOC limit; similar behavior can be observed for the RL policy. Notice that the other control policies manage to smooth the grid power during the day, although there may be slight deviations between the model-based policies and IL and GIL due to the approximate behavior of the NNs.

Fig. 9 depicts the range of occurring voltages, currents, and transformer powers for all policies. Note that for the reference case, SC and RL result in high overloads of more than 120 % of the grid equipment as well as voltage band violations, whereas the other methods lead to a significant reduction in overloads and voltage violations. The model-based methods achieve very good results for the BCF, as they comply with the voltage band and line limits for every time step in the year and only result in a few transformer overloads slightly exceeding 100 %. Voltage band violations and transformer overloads occur with the IL and GIL methods, which perform significantly better than the model-free RL, since RL methods have difficulties in learning the correlations in the grid due to the complexity of the system.

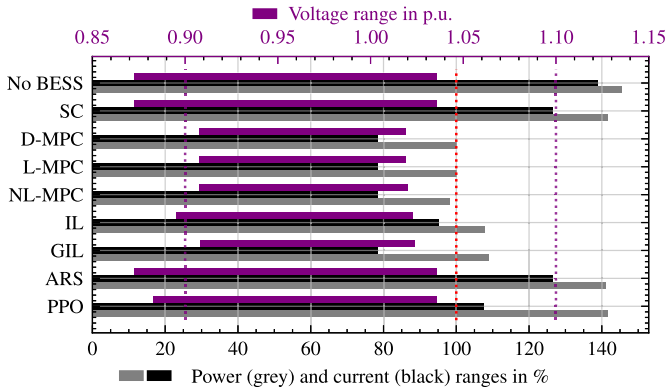


Fig. 9. Range of voltages, currents and transformer power for all policies with BCF.

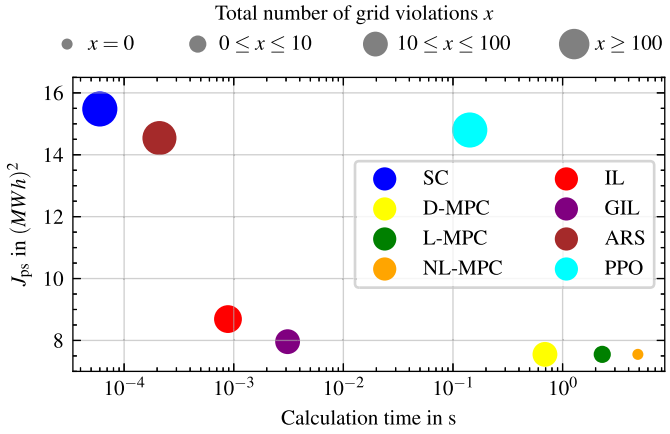


Fig. 10. Objective function values and computation time based on BCF and grid violations.

Fig. 10 illustrates the calculation time of the methods with the corresponding objective function values for BCF from Table 4. The size of the individual points represents the number of grid violations. The trade-off between high model accuracy with few grid violations and long calculation times versus fast calculation times and less accuracy can be seen in this figure. PPO has higher computational time due to the calculation of augmented features for every time step. In the midfield are the model-free approaches from imitation learning, which represent a compromise between model accuracy, grid violations and reasonable computing time. Note that in the considered case, computation times are well below the temporal resolution Δt due to the relatively small grid size and chosen complexity of the building models. It is expected that this might not be the case for larger grids, where the simulation time can exceed the temporal resolution, which would lead to scenarios that cannot be solved in time, making MPC-based approaches impractical. In this case, the model-free approaches can still be applied due to their approximative behavior and significantly lower calculation times.

3.7. Discussion

While the presented study provides valuable insights into the comparison of model-based and model-free control methods for BESS control, it is important to acknowledge several limitations inherent in the analysis.

Thus, first, although a clear trend was observed between the performance of model-based and model-free control methods, revealing trade-offs that highlight the respective strengths and weaknesses of these approaches, the comparison itself was constrained by the scope of the simulation. Specifically, the study was conducted on a single benchmark grid to maintain manageable simulation times for a simulation period of one year with $\Delta t = 15$ min. This restriction inherently limits the diversity of scenarios considered and may not fully capture the performance variations under different grid configurations and/or operating conditions. The general trends observed in this study may be applicable to other similar grids, but to draw definitive conclusions, further analysis of a large number of diverse benchmark grids would be required. In this context, it is important to mention that the obtained results focus on making a comparative investigation of various model-based and model-free HEMS control methods, highlighting their advantages and disadvantages.

Second, the considered case featured a time resolution significantly longer than the calculation times of the model-based methods, which ensured feasibility within the simulation framework. However, extending the analysis to cases where this assumption no longer holds, presents a significant challenge. For instance, considering scenarios where model-free approaches require training data that comprehensively covers all seasons would necessitate simulation times spanning multiple years, making such studies computationally prohibitive.

Third, while the presented results are focused on a simple but practical system, the implications for more complex systems remain uncertain. The inclusion of additional controlled equipment, particularly EV charging and HVAC setpoints, or grids with a larger number of nodes would drastically increase the computational demands for model-based methods, such as MPC. Therefore, it is reasonable to expect a substantial increase in the required calculation times in these cases. In contrast, the model-free approaches benefit from inference times that remain largely independent of problem complexity, offering a potential advantage in larger systems. However, this comes with the caveat that the synthetic data generation for training such methods when actual data is not available would also scale with the complexity of the system.

4. Conclusion

Different methods from the rule-based, model-based and model-free research domains were compared here for a benchmark LV grid to control multiple BESS using the proposed framework. A detailed analysis from a homeowner and DSO perspective was carried out using several metrics, showing that model-based and model-free methods can achieve improvements over a typical, heuristic rule-based controller. The best results in terms of objective function value and constraint violations were achieved with model-based methods, but at the expense of the higher computational times associated with these methods. Model-free approaches resulted in a reduction in computational time of up to four orders of magnitude with a slight loss in accuracy, demonstrating the advantage of these methods for solving complex optimization problems. Furthermore, the effect of different forecast quality available to the methods was investigated, showing an expected decrease in overall performance with lower forecast quality, highlighting the value of using model-free approaches in realistic settings.

Future studies can investigate the application of model-based RL in the presented framework, which may circumvent the limitations of RL highlighted in the paper. Another possible research direction deals with the combination of model-free and model-based methods, taking into account the short computation time of model-free methods and the accuracy and basic physical equations of model-based methods. Furthermore, to prevent poor decisions by RL controllers, hybrid methodologies could be examined, in which a classical heuristic controller may be utilized as a fallback option in such cases. Finally, investigations into applying the proposed methods for grids with a large number of nodes can also be performed to demonstrate the advantages of the reductions in computing time.

CRediT authorship contribution statement

Felicitas Mueller: Visualization, Resources, Investigation, Conceptualization, Writing – review & editing, Validation, Project

administration, Formal analysis, Writing – original draft, Software, Methodology, Data curation. **Steven de Jongh:** Visualization, Resources, Investigation, Conceptualization, Writing – review & editing, Validation, Project administration, Formal analysis, Writing – original draft, Software, Methodology, Data curation. **Claudio A. Cañizares:** Supervision, Conceptualization, Writing – review & editing, Project administration, Visualization, Funding acquisition. **Thomas Leibfried:** Writing – review & editing, Funding acquisition, Supervision, Project administration. **Kankar Bhattacharya:** Writing – review & editing, Supervision, Funding acquisition.

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

During the preparation of this work the authors used AI-based language models DeepL and OpenAI GPT-4 for text editing and language-related tasks. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered potential competing interests:

Felicitas Mueller reports that financial support was provided by Mitacs Inc. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The financial support for this research was provided by Mitacs Canada, as well as the University of Waterloo and Karlsruhe Institute of Technology to conduct research in the field of energy system optimization.

Data availability

Data will be made available on request.

References

[1] Zhou B, Li W, Chan KW, Cao Y, Kuang Y, Liu X, et al. Smart home energy management systems: concept, configurations, and scheduling strategies. Renew Sustain Energy Rev 2016;61:30–40. https://doi.org/10.1016/j.rser.2016.03.047

[2] Ali AO, Elmarghany MR, Abdelsalam MM, Sabry MN, Hamed AM. Closed-loop home energy management system with renewable energy sources in a smart grid: a comprehensive review. J Energy Storage 2022;50:104609. https://doi.org/10.1016/j.est.2022.104609

[3] Beaudin M, Zareipour H. Home energy management systems: a review of modelling and complexity. Renew Sustain Energy Rev 2015;45:318–35. https://doi.org/10.1016/j.rser.2015.01.046

[4] Ma Y, Azuatalam D, Power T, Chapman AC, Verbič G. A novel probabilistic framework to study the impact of photovoltaic-battery systems on low-voltage distribution networks. Appl Energy 2019;254:113669. https://doi.org/10.1016/j.apenergy.2019.113669

[5] Langer L, Volling T. An optimal home energy management system for modulating heat pumps and photovoltaic systems. Appl Energy 2020;278:115661. https://doi.org/10.1016/j.apenergy.2020.115661

[6] Bouakkaz A, Mena AJG, Haddad S, Ferrari ML. Efficient Energy scheduling considering cost reduction and Energy saving in hybrid Energy system with Energy Storage. J Energy Storage 2021;33:101887. https://doi.org/10.1016/j.est.2020.101887

[7] Elkazaz M, Sumner M, Naghiyev E, Pholboon S, Davies R, Thomas D. A hierarchical two-stage energy management for a home microgrid using model predictive and real-time controllers. Appl Energy 2020;269:115118. https://doi.org/10.1016/j.apenergy.2020.115118

[8] Dinh HT, Lee KH, Kim D. Supervised-learning-based hour-ahead demand response for a behavior-based home Energy management system approximating MILP optimization. Appl Energy 2022;321:119382. https://doi.org/10.1016/j.apenergy.2022.119382

[9] Jin X, Baker K, Christensen D, Isley S. Foresee: a user-centric home Energy management system for Energy efficiency and demand response. Appl Energy 2017;205:1583–95. https://doi.org/10.1016/j.apenergy.2017.08.166

[10] Ellis GD, Canizares CA, Bhattacharya K, Bozchalui MC, Hassen H, Hashmi SA. Computer implemented electrical Energy hub management system and method, US Patent 9671843 B2, June 2017, Canadian Patent 2831621C, June 2019.

[11] Mueller F, de Jongh S, Cañizares C, Leibfried T, Bhattacharya K. Impact of predictive and non-predictive battery control methods on residential Buildings and a benchmark distribution grid. Appl Energy 2024. https://doi.org/10.2139/ssrn.4759091

[12] Stoffel P, Maier L, Kümpel A, Schreiber T, Müller D. Evaluation of advanced control strategies for building energy systems. Energy Build 2023;280:112709. https://doi.org/10.1016/j.enbuild.2022.112709

[13] Maier L, Brillert J, Zanetti E, Müller D. Approximating model predictive control strategies for heat pump systems applied to the building optimization testing framework (BOPTEST). J Build Perform Simul 2024;17(3):338–60. https://doi.org/10.1080/19401493.2023.2280577

[14] Lefebure N, Khosravi M, Hudoba de Badyn M, Bünning F, Lygeros J, Jones C, et al. Distributed model predictive control of buildings and energy hubs. Energy Build 2022;259:111806. https://doi.org/10.1016/j.enbuild.2021.111806

[15] Morstyn T, Hredzak B, Agelidis VG. Control strategies for Microgrids with distributed Energy Storage systems: an overview. IEEE Trans. Smart Grid 2018;9(4):3652–66. https://doi.org/10.1109/TSG.2016.2637958

[16] Li P, Wu W, Wang X, Xu B. A data-driven linear optimal Power flow model for distribution networks. IEEE Trans. Power Syst 2023;38(1):956–59. https://doi.org/10.1109/TPWRS.2022.3216161

[17] Canyasse R, Dalal G, Mannor S. Supervised learning for optimal power flow as a real-time proxy. In: IEEE PES ISGT; 2017. p. 1–5. https://doi.org/10.1109/ISGT.2017.8086083

[18] Zamzam AS, Baker K. Learning optimal solutions for extremely fast AC optimal Power flow. In: IEEE SmartGridComm; 2020. p. 1–6. https://doi.org/10.1109/SmartGridComm47815.2020.9303008

[19] de Jongh S, Steinle S, Hlawatsch A, Mueller F, Suriyah M, Leibfried T. Neural predictive control for the optimization of Smart Grid flexibility schedules. In: 56th UPEC; 2021. p. 1–6. https://doi.org/10.1109/UPEC50034.2021.9548179

[20] Li Y, Yu C, Shahidehpour M, Yang T, Zeng Z, Chai T. Deep reinforcement learning for Smart Grid operations: algorithms, applications, and prospects. Proc IEEE 2023;111(9):1055–96. https://doi.org/10.1109/JPROC.2023.3303358

[21] Duan J, Shi D, Diao R, Li H, Wang Z, Zhang B, et al. Deep-reinforcement-learning-based autonomous voltage control for Power Grid operations. IEEE Trans. Power Syst 2020;35(1):814–17. https://doi.org/10.1109/TPWRS.2019.2941134

[22] Wang R, Bu S, Chung CY. Real-time joint regulations of frequency and voltage for TSO-DSO coordination: a deep reinforcement learning-based approach. IEEE Trans. Smart Grid 2024;15(2):2294–308. https://doi.org/10.1109/TSG.2023.3302155

[23] Quakernack L, Kelker M, Haubrock J. Deep reinforcement learning for autonomous control of low voltage Grids with focus on Grid stability in future Power Grids. In: IEEE PES ISGT-Europe; 2022. p. 1–5. https://doi.org/10.1109/ISGT-Europe54678.2022.9960416

[24] Hu J, Ye Y, Tang Y, Strbac G. Towards risk-aware real-time security constrained economic dispatch: a tailored deep reinforcement learning approach. IEEE Trans. Power Syst 2024;39(2):3972–86. https://doi.org/10.1109/TPWRS.2023.3288039

[25] Cao D, Zhao J, Hu J, Pei Y, Huang Q, Chen Z, et al. Physics-informed graphical representation-enabled deep reinforcement learning for robust distribution system voltage control. IEEE Trans. Smart Grid 2024;15(1):233–46. https://doi.org/10.1109/TSG.2023.3267069

[26] de Jongh S, Freund D, Mueller F, Suriyah M, Leibfried T. Reinforcement learning based flexibility optimization in distribution Grids. In: 26th CIRED; 2021. p. 1375–79. https://doi.org/10.1049/icp.2021.2202

[27] Zhang H, Seal S, Wu D, Bouffard F, Boulet B. Building Energy management with reinforcement learning and model predictive control: a survey. IEEE Access 2022;10:27853–62. https://doi.org/10.1109/ACCESS.2022.3156581

[28] Wu H, Pratt A, Munankarmi P, Lunacek M, Balamurugan SP, Liu X, et al. Impact of model predictive control-enabled home energy management on large-scale distribution systems with photovoltaics. Adv Appl Energy 2022;6:100094. https://doi.org/10.1016/j.adapen.2022.100094

[29] Lin X, Zamora R, Jiang Y, Chen G, Srivastava AK. A multi-level Home Energy Management system (HEMS) for DC-Microgrids. IEEE Trans Sustain Energy 2025:1–15. https://doi.org/10.1109/TSTE.2025.3551682

[30] Seal S, Boulet B, Dehkordi VR, Bouffard F, Joos G. Centralized MPC for home Energy management with EV as Mobile Energy Storage unit. IEEE Trans Sustain Energy 2023;14(3):1425–35. https://doi.org/10.1109/TSTE.2023.3235703

[31] Alfaverh F, Denaï M, Sun Y. Demand response strategy based on reinforcement learning and fuzzy reasoning for home Energy management. IEEE Access 2020;8:39310–21. https://doi.org/10.1109/ACCESS.2020.2974286

[32] Zhang C, Kuppannagari SR, Xiong C, Kannan R, Prasanna VK. A cooperative multi-agent deep reinforcement learning framework for real-time residential load scheduling. In: Proceedings of the International Conference on Internet of Things Design and Implementation, IoTDI '19; New York, NY, USA: Association for Computing Machinery; 2019. p. 59–69. https://doi.org/10.1145/3302505.3310069

[33] Moy K, Tae C, Wang Y, Henri G, Bambos N, Rajagopal R. An OpenAI-OpenDSS framework for reinforcement learning on distribution-level microgrids. In: 2021 IEEE Power & Energy Society General Meeting (PESGM); 2021. p. 1–5. https://doi.org/10.1109/PESGM46819.2021.9638106

[34] Amer A, Shaban K, Massoud A. Demand response in HEMSs using DRL and the impact of its various configurations and environmental changes. Energies 2022;15(21):8235. https://doi.org/10.3390/en15218235

[35] Yu L, Xie W, Xie D, Zou Y, Zhang D, Sun Z, et al. Deep reinforcement learning for Smart home Energy management. IEEE Internet Things J 2020;7(4):2751–62. https://doi.org/10.1109/JIOT.2019.2957289

[36] Lee J, Wang W, Niyato D. Demand-side scheduling based on multi-agent deep Actor-critic learning for Smart grids. In: 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids

(SmartGridComm); 2020. p. 1–6. https://doi.org/10.1109/SmartGridComm47815.2020.9302935
[37] Shuvo SS, Yilmaz Y. Home Energy Recommendation system (HERS): a deep reinforcement learning method based on residents' feedback and activity. IEEE Trans Smart Grid 2022;13(4):2812–21. https://doi.org/10.1109/TSG.2022.3158814
[38] Mocanu E, Mocanu DC, Nguyen PH, Liotta A, Webber ME, Gibescu M, et al. On-line building Energy optimization using deep reinforcement learning. IEEE Trans Smart Grid 2019;10(4):3698–708. https://doi.org/10.1109/TSG.2018.2834219
[39] Cai W, Sawant S, Reinhardt D, Rastegarpour S, Gros S. A learning-based model predictive control strategy for home Energy management systems. IEEE Access 2023;11:145264–80. https://doi.org/10.1109/ACCESS.2023.3346324
[40] EnWG §14a. Gesetz über die Elektrizitäts- und Gasversorgung (Energiewirtschaftsgesetz - EnWG) §14a Netzorientierte Steuerung von steuerbaren Verbrauchseinrichtungen und steuerbaren Netzanschlüssen; Festlegungskompetenzen. 11 2023. https://www.gesetze-im-internet.de/enwg_2005/__14a.html [accessed: Oct 11, 2024].
[41] VDE Verband der Elektrotechnik Elektronik Informationstechnik e.V. VDE-AR-N 4105:2018-11: Erzeugungsanlagen am Niederspannungsnetz, Technische Mindestanforderungen für Anschluss und Parallelbetrieb von Erzeugungsanlagen am Niederspannungsnetz [Tech. rep.]. VDE; Nov 2018.
[42] Weniger J, Quaschning V. Begrenzung der Einspeiseleistung von netzgekoppelten Photovoltaiksystemen mit Batteriespeichern. In: Symposium Photovoltaische Solarenergie, Bad Staffelstein; 2013.
[43] Gerards MET, Hurink JL. Robust peak-shaving for a neighborhood with electric vehicles. Energies 2016;9(8). https://doi.org/10.3390/en9080594
[44] Zimmerlin M, Littig D, Held L, Mueller F, Karakus C, Suriyah MR, et al. Optimal and efficient real time coordination of flexibility options in integrated Energy systems. In: International ETG-Congress 2019; ETG Symposium; 2019. p. 1–6.
[45] Bolognani S, Zampieri S. On the existence and linear approximation of the Power flow solution in Power distribution networks. IEEE Trans. Power Syst 2016;31(1):163–72. https://doi.org/10.1109/TPWRS.2015.2395452
[46] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. Adv Neural Inf Process Syst 2017;30. https://doi.org/10.48550/arXiv.1706.02216
[47] de Jongh S, Gielnik F, Mueller F, Schmit L, Suriyah M, Leibfried T. Physics-informed geometric deep learning for inference tasks in power systems. Electric Power Syst Res 2022;211:108362. https://doi.org/10.1016/j.epsr.2022.108362
[48] Mania H, Guy A, Recht B. Simple random search provides a competitive approach to reinforcement learning. arXiv:1803.07055 2018.
[49] Silver D, Lever G, Heess N, Degris T, Wierstra D, Riedmiller M. Deterministic policy gradient algorithms. In: 31st Int. Conference on Machine Learning (ICML); vol. 32. 2014. p. 387–95.
[50] Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. CoRR arXiv:1509.02971. 2017.
[51] Haarnoja T, Zhou A, Abbeel P, Levine S. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: 35th ICML; vol. 80. 2018. p. 1861–70. https://doi.org/10.48550/arXiv.1801.01290
[52] Fujimoto S, van Hoof H, Meger D. Addressing function approximation error in Actor-critic methods. In: 35th ICML; vol. 80. 2018. p. 1587–96. https://doi.org/10.48550/arXiv.1802.09477
[53] Mnih V, Badia AP, Mirza M, Graves A, Lillicrap T, Harley T, et al. Asynchronous methods for deep reinforcement learning. In: 33rd ICML; vol. 48. 2016. p. 1928–37. https://doi.org/10.48550/arXiv.1602.01783
[54] Habbak H, Mahmoud M, Metwally K, Fouda MM, Ibrahem MI. Load forecasting techniques and their applications in Smart Grids. Energies 2023;16(3). https://doi.org/10.3390/en16031480
[55] Haben S, Arora S, Giasemidis G, Voss M, Vukadinović Greetham D. Review of low voltage load forecasting: methods, applications, and recommendations. Appl Energy 2021;304:117798. https://doi.org/10.1016/j.apenergy.2021.117798
[56] Pflugradt ND. Modellierung von Wasser und Energieverbräuchen in Haushalten [PhD dissertation]. TU Chemnitz; 2016.
[57] Orth N, Weniger J, Meissner L. Empfehlungen zur Auslegung von Solarstromspeichern. Sonnenenergie 2022;2:16–17.
[58] Meinecke S, Sarajlić D, Drauz SR, Klettke A, Lauven L-P, Rehtanz C, et al. Simbench—A benchmark dataset of electric Power systems to compare innovative solutions based on Power flow analysis. Energies 2020;13(12). https://doi.org/10.3390/en13123290
[59] Thurner L, Scheidler A, Schafer F, Menke JH, Dollichon J, Meier F, et al. Pandapower - an open source Python tool for convenient modeling, analysis and optimization of electric Power systems. IEEE Trans. Power Syst 2018. https://doi.org/10.1109/TPWRS.2018.2829021
[60] Hart WE, Watson J-P, Woodruff DL. Pyomo: modeling and solving mathematical programs in Python. Math Program Comput 2011;3(3):219–60. https://doi.org/10.1007/s12532-011-0026-8
[61] Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. Automatic differentiation in PyTorch. In: 3con1st NIPS Conference; 2017.
[62] Fey M, Lenssen JE. Fast graph representation learning with PyTorch Geometric. arXiv:1903.02428 2019.
[63] Raffin A, Hill A, Gleave A, Kanervisto A, Ernestus M, Dormann N. Stable-Baselines3: reliable reinforcement learning implementations. J Mach Learn Res 2021;22(268):1–8.
[64] Munzke N, Schwarz B, Hiller M. Evaluation of the efficiency and resulting electrical and economic losses of photovoltaic home storage systems. J Energy Storage 2021;33:101724. https://doi.org/10.1016/j.est.2020.101724
[65] Jasper FB, Späthe J, Baumann M, Peters JF, Ruhland J, Weil M. Life Cycle Assessment (LCA) of a battery home Storage system based on primary data. J Cleaner Prod 2022;366:132899. https://doi.org/10.1016/j.jclepro.2022.132899
[66] Xu S, Zhu J, Li B, Yu L, Zhu X, Jia H, et al. Real-time power system dispatch scheme using grid expert strategy-based imitation learning. Int J Electr Power Energy Syst 2024;161:110148. https://doi.org/10.1016/j.ijepes.2024.110148
[67] Liu M, Guo M, Fu Y, O'Neill Z, Gao Y. Expert-guided imitation learning for energy management: evaluating gail's performance in building control applications. Appl Energy 2024;372:123753. https://doi.org/10.1016/j.apenergy.2024.123753
[68] Dongol D, Feldmann T, Schmidt M, Bollin E. A model predictive control based peak shaving application of battery for a household with photovoltaic system in a rural distribution grid. Sustain Energy Grids Netw 2018;16:1–13. https://doi.org/10.1016/j.segan.2018.05.001