Optimization of demand response-oriented electrolytic and fuel cell cogeneration system for community residents: uncovering flexibility and gaps

Optimization of demand response-oriented electrolytic and fuel cell cogeneration system for community residents: uncovering flexibility and gaps

Xiaohai Zhang[1], José Luis Ramírez-Mendiola[2], Yongquan Lai[1], Jinzhan Su[1], Mingtao Li[1*], Liejin Guo[1]
[1] International Research Center for Renewable Energy (IRCRE), State Key Laboratory of Multiphase Flow in Power Engineering (MFPE), Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China.
[2] School of the Built Environment, University of Reading, Whiteknights, Reading RG6 6AF, United Kingdom.

[*] Corresponding author.
E-mail address: mingtao@xjtu.edu.cn

Abstract

Low carbon energy systems are dependent on renewable power sources, which present challenges in controllability compared to conventional sources. This poses difficulties in maintaining grid balance. To address these challenges, demand response mechanisms and low-carbon technologies are being implemented, particularly in the residential sector, which is a significant consumer of electricity and heat. On-site hydrogen production by alkaline electrolytic cells (AEC) and proton exchange membrane fuel cell - combined heat and power (PEMFC-CHP) systems are of particular interest for their potential to improve grid flexibility. This study compares the performance of an AEC-PEMFC-CHP system with a market-competitive heat pump (HP) system for a residential community scenario. Results from a two-step capacity-operation optimization using advanced methods demonstrate that the AEC-PEMFC-CHP system offers greater flexibility but at the expense of higher power consumption and lower efficiency. Additionally, the cost of the AEC-PEMFC-CHP system is 2.3 times higher than that of the typical HP system, implying that it would require a 0.6 times higher efficiency and a 1/10th lower equipment cost to compete. Furthermore, load prediction plays a critical role in optimizing both systems, with a longer prediction horizon of 16-20 hours proving effective even with larger prediction errors.

Keywords: Demand response; Optimization; Load uncertainty; Fuel cell; Electrolytic cell

Nomenclature

AEC	Alkaline electrolytic cell
CA	Canlendar aging
COP	Coefficient of performance
CHP	Combined heat and power
DR	Demand response
DSM	Demand-side management
DHW	Domestic hot water
dToU	Dynamic Time-of-Use
EL	Electricity load
EC	Electrolytic cell
FC	Fuel cell
HL	Heat load
HP	Heat pump
HWT	Hot water tank
HT	Hydrogen tank
LHS	Loss of heat supplied
LHSP	Probability of loss of heat supplied
LSTM	Long-short term memory
MPC	Model predictive control
MLP	Multilayer perceptron
NLP	Nonlinear programming
OCV	Open circuit operation
PSO	Particle swarm optimization
PEMFC	Proton exchange membrane fuel cell
QLSTM	Quantile long-short term memory
QR	Quantile regression
RL	Reinforcement learning
RE	Renewable energy
SAC	Soft actor-critic
SH	Space heating
SSC	Start-stop cycles
SoC	State of capacity
SP	Stochastic programming
VC	Voltage cycles

1. Introduction

Several major economies have pledged to reach carbon neutrality by 2050 [1]. However, in order to meet the ambitious decarbonisation goals pledged, the supply of both electricity and heat will need to rely

almost entirely on low-carbon generation technologies [2]. The decarbonisation through electrification paradigm has become mainstream largely due to considerations around energy security and future low-carbon development [1]. However, the unprecedented transformation of the energy systems needed will see them abandon the conventional, highly-controllable, carbon-intensive sources in favour of less-controllable, clean power generation sources. This associated loss of control over the generation output of the system's power sources is problematic, as the approach to grid balancing to this date has been largely based on a supply-follows-demand model. As the shares of less-controllable power sources increase, so do the challenges associated with keeping the grid in balance.

There are two main approaches to addressing these challenges. On the one hand, we can decouple the timing of power consumption with actual demand; on the other hand, we can directly shift the demand so that it matches available supply [3]. Decoupling strategies are mostly based on the use of energy storage resources, which allow for passively shifting energy demand in time and space to avoid mismatches between supply and demand. Direct matching strategies rely on the use of flexible power sources in combination with demand-side management (DSM) programmes; thus, supply and demand are matched by adjusting either power supply or demand load, or both. In particular, DSM has received a lot of attention because it has the potential to reduce or entirely avoid additional investments in infrastructure, making the corresponding flexibility capital increasingly important [4].

When it comes to matching strategies, there is already a wealth of practice in optimizing flexibility on the supply side, but the flexibility on the demand side brought by digitization has vast potential for exploitation [5]. In this regard, demand response (DR) in industry and commerce has been proven in practice to provide reliable flexibility in the grid, while the rapidly developing electric vehicles are seen as the most promising source of mobile power in the transport sector, and flexibility in the residential sector is expected to be promoted by heat pumps and energy storage systems [6]. Morever, adjustments from household electricity consumption behaviour can provide 1.9-3.9% of flexibility [7]. Taking into account that the residential sector is one of the primary consumers of electricity and heat (close to 1/3 in developed economies) and that their energy demand is highly intrinsically similar, flexibility solutions for the residential sector have significant scalability benefits [8-10].

In combination with the electrification of heating, energy storage and demand response, heat pump systems are certainly a promising solution for flexible energy usage, as has been proven in practice [11, 12]. However, heat pumps offer a one-way conversion of electricity to heat for the purpose of consuming renewable energy (RE) power in surplus, but they cannot supply the power needed for the load when RE power is insufficient. In contrast, the use of electrolytic cells (EC) for the production of hydrogen, bundled together with fuel cell based combined heat and power system (FC-CHP), has the potential to allow for a seamless decoupling of the timing of power demand and actual consumption. The application of the Power-to-Hydrogen strategy has attracted a great deal of interest [13], where the intermediate hydrogen, as a carbon-free fuel, is considered to hold the promise of creating a sustainable and emission-free energy

cycle to integrate more renewable energy sources [2, 14, 15].

Although the implementation of many FC-CHP projects, such as Ene-Farm, Ene Field and Callux, has led to significant improvements in costs and performance, the economic competitiveness of FC-CHP still needs to be further improved. Starting with a system optimisation perspective, Adam et al. [16] highlight the importance of matching between heat production and load profiles to optimize the design of FC-CHP system. And practice from the Italian project MICROGEN 30 reveals that scaling up the load and considering the power sales benefits are important to improve the economics of the proton exchange membrane fuel cell (PEMFC) based CHP system [17]. Movever, with the development of inverse metering and electrical demand response, more advanced operating strategies are required to deliver better performance throughout the annual load curve, while offering high efficiency and low fuel consumption [18].

Traditional operating strategies of FC-CHP system are based on passive electricity/heat load following. Compared to typical heat-led strategy for a system without heat storage, Laurens et al. [19] point out the electricity-led is more cost-competitive in Germany, and thus studies a PEMFC-CHP system with auxiliary boiler and electric storage to explore its economic feasibility under a fixed electricity tariff. The results demonstrate that PEMFC-CHP requires considerable time to achieve a substantial market diffusion, and the electric storage is not a worthwhile extension for a grid connected scenario. When more factors are taken into account, Mosayeb et al. [20] propose a modified firefly algorithm to optimize stochastic scheduling of PEMFC-CHP in a hybrid energy system including PV, WT and thermal power plant, on a 33-bus distributed network, in which the market profit, emission and average energy not support are taken as the objectives optimized based on the generated scenarios considering the uncertainty from wind speed, solar radiation and electricity market price. Furthermore, considering the interaction between different agents, Dinh et al. [21] propose a peer-to-peer (P2P) electricity trading system designed for optimal energy management in a group of dwellings to help operate FC-CHP systems at a higher efficiency and increase the self-sufficiency of such dwellings. In general, due to the high cost and complexity of FC-CHP systems, careful sizing and advanced integration strategies must be employed.

The degree to which applications for these systems can be found is determined not only in terms of capacity but also, and more importantly, in terms of operating modes and strategies. Therefore, advanced optimization methods are playing an increasingly important role in improving the economics and performance of FC-CHP systems in the face of increasingly complex system environments. This is especially evident when the hydrogen is sourced from electricity rather than from the independently supplied natural gas used in the vast majority of studies to date, which places greater demands on system optimization. In addition, consideration of dynamic or quasi-steady-state system modelling supplemented by measured data and practices from existing commercial units, and taking into account economic parameters and degradation factors in advanced optimization methods on realistic load profiles, are essential for a more realistic assessment of fuel cell-based CHP systems [22].

In response to the observed trend outlined above, this paper presents novel contributions to the energy research literature, namely:

- It's the first systematic analysis of grid-based EC-FC-CHP systems for residential community electricity and heat, taking into account FC degradation and demand response-oriented optimization, clearly measuring the technical and economic gaps compared to HP systems;
- the flexibility and impacts of grid-based HP system and EC-FC-CHP system for community residents on the power grid have been systematically analyzed between different operation strategies;
- the impacts of load uncertainty represented by a quantile long-short term memory (QLSTM) neural network model on the operation optimization have been fully explored; and
- comparisons between swarm intelligence and data-driven optimization methods for capacity design, and between model predictive control (MPC) and reinforcement learning (RL) methods for operation optimization have been carried out to assess their characteristics;

To demonstrate the technical and economic gap, grid-based HP systems are used as Baseline Scenario (S_B) and grid-based EC-FC-CHP systems are referred to as Cogeneration Scenario (S_C). In general, the S_B scenario, which contains heat pump (HP) and hot water tank (HWT), can be considered as a short-term, market-competitive solution for limited grid balancing. While the S_C scenario, which contains alkaline electrolytic cell (AEC), hydrogen tank (HT), proton exchange membrane fuel cell based combined heat and power (PEMFC-CHP) and HWT, can be considered as a long-term solution with greater flexibility, but currently less cost-competitive. Both scenarios are capacity- and operation-optimized to achieve optimal performance; particle swarm optimization (PSO) in swarm intelligence and Bayesian optimization (BO) in data-driven methods are used for capacity optimization following a heat-led strategy; MPC based on nonlinear programming (NLP) and stochastic programming (SP), and the RL algorithm named soft actor-critic (SAC), designed for continuous space control, are implemented based on the optimal capacity results to optimize the cost, load leveling and PEMFC lifetime. In terms of electricity and heat load profiles, a sample community consisting of 242 households in the London area provides data for the whole year of 2013; as part of a DR trial, these households were notified of the dynamic time-of-use (dToU) tariffs to be enforced the following day. During the capacity optimization stage, the true loads are used, while in operation optimization, predictive loads from point prediction without uncertainty information and interval prediction with uncertainty information are delivered under different prediction horizons, on a rolling basis. In addition, the PEMFC model considering cell voltage degradation and partial load efficiency is proposed for a more realistic assessment. The workflow of the research presented in this paper is summarised in Figure 1.

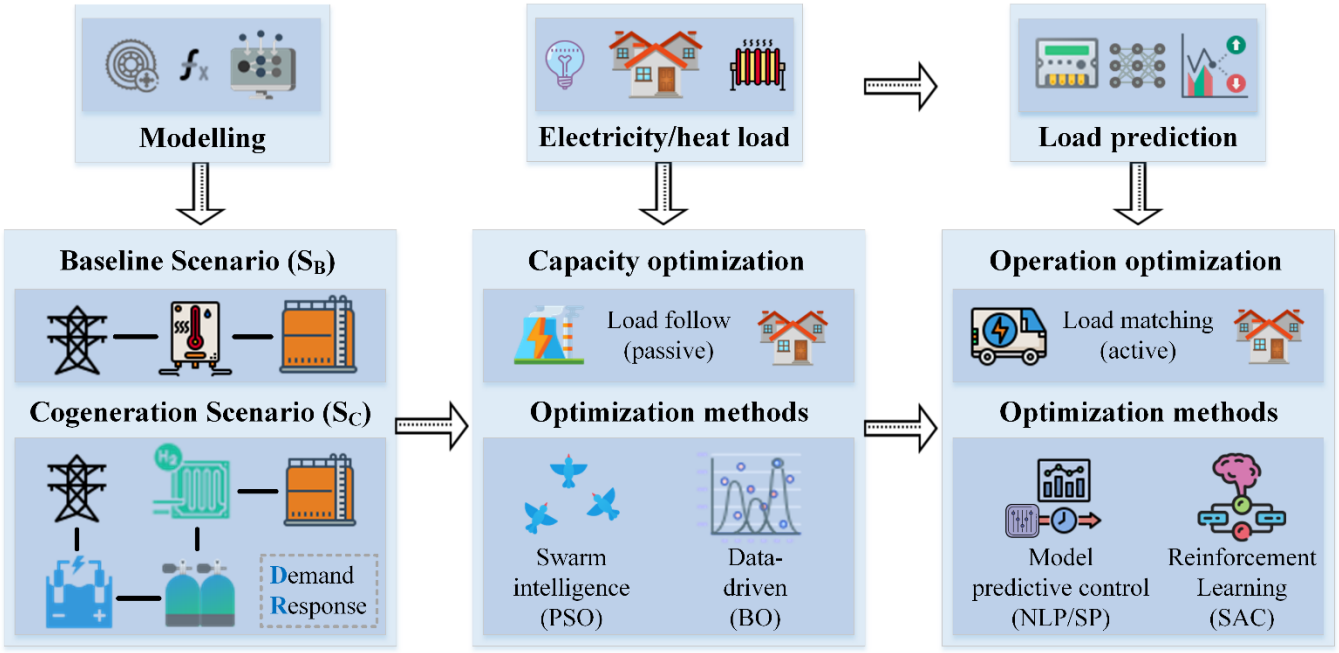


Figure 1 – Workflow of this research project.

What remains of this paper is arranged as follows: Section 2 introduces the problem formulation; Section 3 describes the modelling of components; Section 4 illustrates the modelling of load; Section 5 demonstrates the optimization methods; Section 6 presents the calculation results and relative discussion; finally, Section 7 offers our concluding remarks.

2. Problem formulation

For offering flexibility to grid, the system structure and energy flows of the technologies considered in HP based S_B and hydrogen based S_C are shown in Figure 2. In S_B, there is only a single conversion from electricity to heat, where the electricity load is met by grid electricity and the heat load is supplied by the HP and HWT. On the other hand, in S_C, bidirectional conversion of electricity-hydrogen-electricity/heat is achieved, where the electricity load is satisfied by PEMFC and grid, and the heat load is provided by PEMFC and HWT. In addition, the PEMFC is able to export electricity to the grid as a distributed power source.

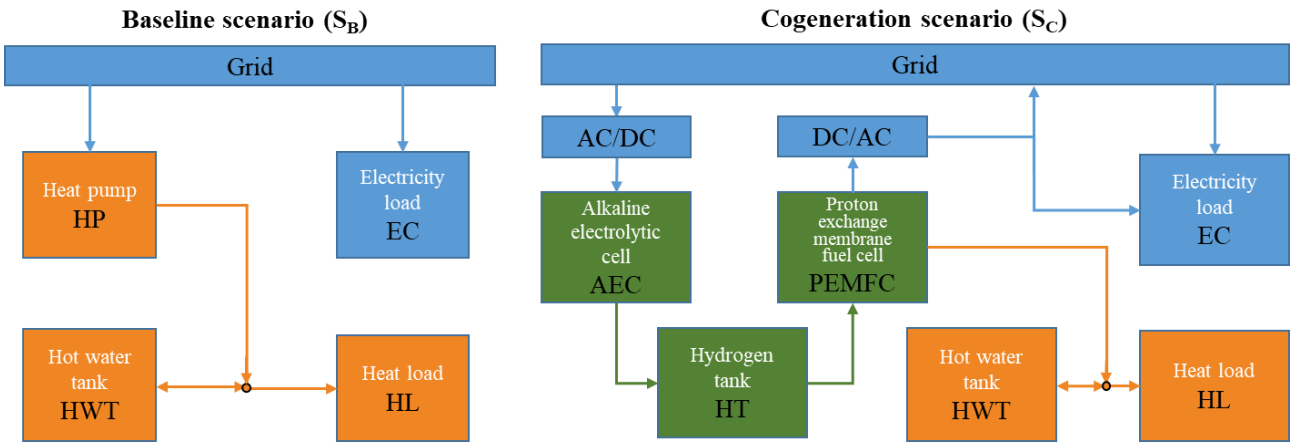


Figure 2 – System structure and energy flow of baseline (S_B) and cogeneration (S_C) scenarios.

In order to sufficiently improve the cost and load leveling of proposed systems, the optimization process has been divided into two steps: capacity and operation. In capacity optimization, the cost is the only objective with the constraints of reliability and physical mechanism, following a heat-led strategy. Then, based on the optimal solution of capacity optimization, the operation optimization aims to minimize the cost and improve the load leveling and life of PEMFC.

2.1 Capacity optimization

For capacity optimization purposes, the rated capacity of components will be optimized according to specific objectives and constraints. The optimization variables have been summarized in Table 1, where the $\mathrm{P}, \mathrm{Q}, \mathrm{m}$ denote the power (kW), quantity of heat (kWh) and mass (kg), respectively. In the S_B using heat pump, the S_{B0} represents a solution without thermal storage (i.e. HWT), while the S_{B1} denotes a solution with thermal storage. In the S_C, the S_{C1} denotes a solution with optimal capacities for a hydrogen-based AEC-HT-PEMFC-CHP-HWT system.

Table 1 – Optimization variables in baseline (S_B) and cogeneration (S_C) scenarios for capacity optimization.

Variable	Value set (discrete type)	S_{B0}	S_{B1}	S_{C1}
$P_{\mathrm{HP}}^{\mathrm{rated}}$ (kW)	$\{0,10,\cdots,1600\}$	√	√	
$Q_{\mathrm{HWT}}^{\mathrm{rated}}$ (kWh)	$\{0,10,\cdots,33000\}$		√	√
$SoC_{\mathrm{HWT,DHW}}^{\mathrm{thre}}$, $SoC_{\mathrm{HWT,SH}}^{\mathrm{thre}}$	$\{0,0.05,\cdots,1.00\}$		√	√
$P_{\mathrm{el,PEMFC}}^{\mathrm{rated}}$ (kW)	$\{0,10,\cdots,1000\}$			√
$P_{\mathrm{AEC}}^{\mathrm{rated}}$ (kW)	$\{0,10,\cdots,3500\}$			√
$m_{\mathrm{HT}}^{\mathrm{rated}}$ (kg)	$\{0,10,\cdots,1700\}$			√
$SoC_{\mathrm{HT,DHW}}^{\mathrm{thre}}$, $SoC_{\mathrm{HT,SH}}^{\mathrm{thre}}$	$\{0,0.05,\cdots,1.00\}$			√

Notes: the upper boundaries of value sets are estimated according to the mean or maximum heat load.

The heat load following strategy is used during the calculation, in which the thresholds in heating season $SoC_{i,\mathrm{SH}}^{\mathrm{thre}}$ and non-heating season $SoC_{i,\mathrm{DHW}}^{\mathrm{thre}}$ have been introduced to control whether the storage device should be charged, since the mean heat load of heating season mainly for space heating (SP) from October to May of the following year is over 5 times that of non-heating season mainly for domestic hot water (DHW). When the state of capacity (SoC) of the storage device is less than its threshold $\mathrm{SoC}_i^{\mathrm{thre}}$, then the storage device needs to be charged if the source device is available.

As there are no carbon emission during the operation of S_B and S_C, only the economic objective cost of energy (COE) is optimized, which is expressed as follows:

$$\mathrm{COE} = \frac{\mathrm{CRF_{sys}} \times \text{Initial investment cost} + \text{O\&M cost} + \mathrm{CRF_{re}} \times \text{Replacement cost}}{\text{Annual electricity load} + \text{Annual heat load satisfied}} \quad (1)$$

where CRF denotes the capital recovery factor and can be calculated using the interest rate r and targeted number of year n, as follows:

$$\mathrm{CRF} = \frac{r(1+r)^n}{(1+r)^n - 1} \quad (2)$$

For the system's $\mathrm{CRF_{sys}}$, r takes the annual average real interest rate of 2021 in Eurozone which is -1.3%, according to the European Central Bank data; n is the system design lifetime $\mathrm{n_{sys}^{life}}$.
The initial investment cost includes equipment costs and electricity consumption ($\mathrm{W_e}$) cost for initial energy storage, which is expressed as:

$$\text{initial investment cost} = \sum_i \mathrm{cost}_i^{\mathrm{ini}} capacity_i^{\mathrm{rated}} + \sum_i \mathrm{c_e^{normal}} \mathrm{W}_{i,\mathrm{e}}^{\mathrm{ini}} \quad (3)$$

The operation and maintenance (O&M) costs consists of electricity consumption cost and maintenance cost during operation, which is represented using the calculation results of 2013 year, as follow:

$$\text{O\&M cost} = \sum_i \mathrm{cost}_i^{\mathrm{ma}} capacity_i^{\mathrm{rated}} + \sum_t \mathrm{c_e^t W_e^t} \quad (4)$$

There is only replacement cost for stack of PEMFC-CHP module, which is calculated based on the collection of replacement years, as follows:

$$\mathrm{CRF_{re}} \times \text{Replacement cost} = \sum_{t_{\mathrm{re}}} \frac{r(1+r)^{n_{t\mathrm{re}}}}{(1+r)^{n_{t\mathrm{re}}} - 1} \mathrm{cost}_{\mathrm{PEMFC}}^{\mathrm{re}} P_{\mathrm{el,PEMFC}}^{\mathrm{rated}} \quad (5)$$

In the denominator of the previous expression, the annual electricity load is fully supported since it is a

grid-based system. And the annual heat load satisfied $Q_{HL}^{satisfied}$ depends on the loss of heat supplied (LHS) Q_{LHS}^{annual}, which is calculated as:

$$Q_{HL}^{satisfied} = Q_{HL}^{annual} - Q_{LHS}^{annual} = \sum_t Q_{HL}^t - \sum_t Q_{LHS}^t \quad (6)$$

In addition to the objective function, the constraints are built according to the request of reliability and energy balance, as follows:

$$LHSP = \frac{Q_{LHS}^{annual}}{Q_{HL}^{annual}} \leq 0.01 \quad (7)$$

$$0 \leq SoC_i \leq 1 \quad (8)$$

$$SoC_{HWT}^{t+1} = SoC_{HWT}^t + \frac{\Delta Q_{HWT}}{Q_{HWT}^{rated}}, \text{ with } SoC_{HWT}^{ini} = SoC_{HWT,DHW}^{thre} \quad (9)$$

$$SoC_{HT}^{t+1} = SoC_{HT}^t + \frac{\Delta m_{HT}}{m_{HT}^{rated}}, \text{ with } SoC_{HT}^{ini} = SoC_{HT,DHW}^{thre} \quad (10)$$

The LHSP corresponds to the probability of loss of heat supplied, which has been limited to [0,0.01]. When the solution violates this constraint, the absolute value of the violated part is multiplied by 1000 and added to the objective function. For the sake of consistency, energy balances of electricity and heat between different components and loads have been carried out in the simulation of system operation.

2.2 Operation optimization

In contrast to the passive heat load following strategy for operation in capacity design, this part focuses on active load matching operation referring to economic cost and flexibility for electrical peak and trough adjustment, based on the optimal solutions of capacity optimization. In the following, the solutions using model predictive control are denoted by S_{B2} and S_{C2}, while solutions using reinforcement learning are denoted by S_{B3} and S_{C3}.

2.2.1 Model predictive control

The variables for operation optimization using MPC can be divided into controlled variables and noncontrolled variables, which have been listed in Table 2. It should be noted that when the algorithm uses loads from interval prediction, the noncontrolled variables will be accordingly added with respect to three predictive load sceneries with quantile $q = \{0.05, 0.5, 0.95\}$ of load, respectively.

Table 2 – Optimization variables of operation optimization using MPC.

Variable ($j = \{1, \dots, T\}$)	Value range (continuous type)	S_{B2}	S_{C2}

Controlled variables	P_{HP}^{t+j} (kW)	$[0, \mathrm{P}_{\mathrm{HP}}^{\mathrm{rated}}]$	√	
	$P_{\mathrm{el,PEMFC}}^{t+j}$ (kW)	$[0, \widehat{\mathrm{P}}_{\mathrm{el,PEMFC}}^{\mathrm{rated}}]$		√
	P_{AEC}^{t+j} (kW)	$[0, \mathrm{P}_{\mathrm{AEC}}^{\mathrm{rated}}]$		√
Noncontrolled variable	$Q_{\mathrm{HWT,charge}}^{t+j}$ (kWh)	$(0, \mathrm{Q}_{\mathrm{HWT}}^{\mathrm{rated}})$	√	√
	$Q_{\mathrm{HWT,discharge}}^{t+j}$ (kWh)	$(0, \mathrm{Q}_{\mathrm{HWT}}^{\mathrm{rated}})$	√	√
	SoC_{HWT}^{t+j}	[0,1]	√	√
	$m_{\mathrm{HT,charge}}^{t+j}$ (kg)	$[0, \mathrm{m}_{\mathrm{HT}}^{\mathrm{rated}}]$		√
	$m_{\mathrm{HT,discharge}}^{t+j}$ (kg)	$[0, \mathrm{m}_{\mathrm{HT}}^{\mathrm{rated}}]$		√
	SoC_{HT}^{t+j}	[0,1]		√
	States of PEMFC ($U_{\mathrm{cell}}^{t+1}, i^{t+1}, \dot{m}_{\mathrm{H_2}}^{t+1}, \Delta U_{\mathrm{total}}^{t+1}$)			√

Notes: there are only state variables of PEMFC at next time step due to the tradeoff between the solving speed and effectness, which corresponds to one time step optimization of voltage loss for PEMFC. The detailed explanation of variables can be found in Section 3.

The objective of MPC optimization problems includes several parts referring to cost term, load term and PEMFC lifetime term. In the cost term, it aims to reduce total fuel cost, which is related to source and storage components. The load term is designed to make the source components respond to electrical peaks and troughs, in which the source components are supposed to decrease electricity consumption during peak but to increase consumption at trough periods. The thresholds defining the peak and trough correspond to 95% (39.37 kWh) and 5% (10.85 kWh) of electricity load at time resolution of 15 mins. Finally, the PEMFC lifetime term helps to reduce the voltage loss of operation for PEMFC in S_C. In addition, some constraint functions have been added to guarantee reasonable running considering reliability, energy balance and physical mechanism.

In general, following minimization form, the objective function of S_{B2} and S_{C2} are expressed as follows:

$$\mathrm{J}_{\mathrm{S_{B2}}} = \mathrm{J}_{\mathrm{source}}(P_{\mathrm{HP}}) + \mathrm{J}_{\mathrm{storage}}(Q_{\mathrm{HWT,discharge}}, Q_{\mathrm{HWT,charge}}) - \mathrm{J}_{\mathrm{load}}(P_{\mathrm{HP}}) \quad (11)$$

$$\mathrm{J}_{\mathrm{S_{C2}}} = \mathrm{J}_{\mathrm{source}}(P_{\mathrm{AEC}}, m_{\mathrm{HT,charge}}, P_{\mathrm{el,PEMFC}}) + \mathrm{J}_{\mathrm{storage}}(m_{\mathrm{HT,discharge}}, Q_{\mathrm{HWT,discharge}}) - \mathrm{J}_{\mathrm{load}}(P_{\mathrm{el,PEMEC}}, P_{\mathrm{AEC}}) + \mathrm{J}_{\mathrm{life}}(P_{\mathrm{el,PEMFC}}) \quad (12)$$

While using predictive loads from interval prediction, the stochastic programming (SP), considering three scenarios with respect to quantiles $\mathrm{q} = \{0.05, 0.5, 0.95\}$ of load, will be used by increasing the constraints accordingly and producing a weighted objective function corresponding to weights $\mathrm{w}_q = \{0.05, 0.9, 0.05\}$, as follows:

$$\mathrm{J}_{\mathrm{S_{B2}}}^{\mathrm{q}} = \mathrm{J}_{\mathrm{source}}(P_{\mathrm{HP}}) + \sum_q \mathrm{w}_q \, \mathrm{J}_{\mathrm{storage}}^{q}(Q_{\mathrm{HWT,discharge}}, Q_{\mathrm{HWT,charge}}) - \sum_q \mathrm{w}_q \, \mathrm{J}_{\mathrm{load}}^{q}(\mathrm{P}_{\mathrm{HP}}) \quad (13)$$

$$\mathrm{J}_{\mathrm{S_{C2}}}^{\mathrm{q}} = \sum_q \mathrm{w}_q \, \mathrm{J}_{\mathrm{source}}^{q}(P_{\mathrm{AEC}}, m_{\mathrm{HT,charge}}, P_{\mathrm{el,PEMFC}}) + \sum_q \mathrm{w}_q \, \mathrm{J}_{\mathrm{storage}}^{q}(m_{\mathrm{HT,discharge}}, Q_{\mathrm{HWT,discharge}}) - \sum_q \mathrm{w}_q \, \mathrm{J}_{\mathrm{load}}^{q}(P_{\mathrm{el,PEMEC}}, P_{\mathrm{AEC}}) + \mathrm{J}_{\mathrm{life}}(P_{\mathrm{el,PEMFC}}) \quad (14)$$

Detailed descriptions of objective and constraint functions of S_{B2} and S_{C2} have been provided in Appendix B.1.

2.2.2 Reinforcement learning

In contrast to MPC, the RL has the same controlled variables which are taken as the action space listed in Table 3. While the controlled variables only contain actions at the next time step, regardless of the length of predictive loads.

Table 3 – Optimization variables of operation optimization using RL.

Scenario	Action space (controlled variable) (kW)	Value range (continuous type)
S_{B3}	$[P_{\mathrm{HP}}^{t+1}] \in \mathbb{R}^{1\times1}$	$P_{\mathrm{HP}}^{t+1} \in [0, \mathrm{P}_{\mathrm{HP}}^{\mathrm{rated}}]$
S_{C3}	$[P_{\mathrm{el,PEMFC}}^{t+1}, P_{\mathrm{AEC}}^{t+1}] \in \mathbb{R}^{1\times2}$	$P_{\mathrm{el,PEMFC}}^{t+1} \in [0, \mathrm{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated}}]$; $P_{\mathrm{AEC}}^{t+1} \in [0, \mathrm{P}_{\mathrm{AEC}}^{\mathrm{rated}}]$

The training of RL aims to maximize the reward which commonly consists of positive and negative feedbacks. Therefore, the opposite of the MPC objective functions ($-\mathrm{J}_{\mathrm{S_{B2}}}$ and $-\mathrm{J}_{\mathrm{S_{C2}}}$) are used to build the reward function. Besides, the constraint of LHSP is achieved by adding positive feedback for $\mathrm{LHSP}^{t+1} \leq 0.01$ and negative feedback for $\mathrm{LHSP}^{t+1} > 0.01$ to the reward. And other constraints have been realized in the simulation of system operation.

In general, the reward function is the same for loads from point and interval prediction in S_{B3} and S_{C3}, which are expressed as follows:

$$\mathrm{R}_{\mathrm{S_{B3}}} = \mathrm{R}_{\mathrm{LHSP}}^{\mathrm{S_{B3}}}(\mathrm{LHSP}^{t+1}) - \mathrm{J}_{\mathrm{S_{B2}}} \quad (15)$$

$$\mathrm{R}_{\mathrm{S_{C3}}} = \mathrm{R}_{\mathrm{LHSP}}^{\mathrm{S_{C3}}}(\mathrm{LHSP}^{t+1}) - \mathrm{J}_{\mathrm{S_{C2}}} \quad (16)$$

where

$$\mathrm{R}_{\mathrm{LHSP}}^{\mathrm{S_{B3}}}(\mathrm{LHSP}^{t+1}) = \begin{cases} 1, & \mathrm{LHSP}^{t+1} \leq 0.01 \\ -\mathrm{LHS}^{t+1}, & \mathrm{LHSP}^{t+1} > 0.01 \end{cases} \quad (17)$$

$$\mathrm{R}_{\mathrm{LHSP}}^{\mathrm{S_{C3}}}(\mathrm{LHSP}^{t+1}) = \begin{cases} 100, & \mathrm{LHSP}^{t+1} \leq 0.01 \\ -5\mathrm{LHS}^{t+1}, & \mathrm{LHSP}^{t+1} > 0.01 \end{cases} \quad (18)$$

The value of $R_{LHSP}^{S_{B3}}$ and $R_{LHSP}^{S_{C3}}$ are designed to be of a similar order of magnitude to the other terms and have been adjusted in practice.

3. Modelling of components

This work concentrates on energy flow optimization, therefore a balance is struck between accuracy and computational efficiency of the component models. In building PEMFC model, a semi-experienced method for evaluating its voltage loss is used, as the lifetime of PEMFC is relatively short and highly depends on its voltage loss. Besides, an equation evaluating the partial load efficiency of both electrical and thermal output of PEMFC is provided from real world units. Finally, it is noted that the AEC and PEMFC are designed to work simultaneously by using two hydrogen tanks. In the following, the capacity of the hydrogen tanks (m_{HT}) represents the total capacity of the hydrogen storage system.

3.1 Heat pump module

In heating mode, an air source heat pump (HP) driven by power is able to extract heat from ambient air and then transfer the heat into heating zone, following the reverse Carnot cycle principle. The ratio between the heat energy produced and electric energy consumed is called coefficient of performance (COP). In practice, the real COP is used to calculate the heat output for a given power, as follow:

$$\mathrm{COP}_{\mathrm{real}} = \frac{\text{Heat produced } Q_{\mathrm{HP}}}{\text{Electricity consumed } W_{\mathrm{HP}}} = \eta_{\mathrm{Lorenz}} \mathrm{COP}_{\mathrm{Lorenz}} = \eta_{\mathrm{Lorenz}} \frac{T_{\mathrm{lm,sink}}}{T_{\mathrm{lm,sink}} - T_{\mathrm{lm,source}}} \quad (19)$$

where η_{Lorenz} denotes the Lorenz-efficiency of the heat pump. T_{lm} is the log mean temperature of the source and sink heat exchangers, which is calculated as:

$$T_{\mathrm{lm}} = \frac{T_{\mathrm{in}} - T_{\mathrm{out}}}{\ln \frac{T_{\mathrm{in}}}{T_{\mathrm{out}}}} \quad (20)$$

where T_{in} and T_{out} denote the inlet and outlet temperature of the heat exchanger.

3.2 Hot water tank module

A hot water tank (HWT) is a kind of flexible and cheap heat storage solution, and thus has been widely used in residential heating systems. For the purposes of energy flow optimization, only the general heat loss rate per day, $\eta_{\mathrm{loss,\,HWT}}$ of 0.2%, for the hot water tank is considered [23].

3.3 Alkaline electrolytic cell module

In order to produce hydrogen from electricity, the alkaline electrolytic cell (AEC) is selected as it is currently the most commercially available technology, in spite of its relatively low efficiency, η_{AEC} of 0.665 (calculated by lower heating value (LHV) of hydrogen) [24].
Given the power of AEC (P_{AEC}), the mass of hydrogen ($m_{\mathrm{H_2,AEC}}$) produced in Δt hour is calculated by:

$$m_{\mathrm{H_2,AEC}} = \frac{3600 P_{\mathrm{AEC}} \Delta t}{\mathrm{LHV_{H_2}/M_{H_2}}} \eta_{\mathrm{AEC}} \quad (21)$$

where $\mathrm{LHV_{H_2}}$ and $\mathrm{M_{H_2}}$ denote the lower heating value (241 $\mathrm{kJ \cdot mol^{-1}}$) and molar mass (2 $\mathrm{g \cdot mol^{-1}}$) of hydrogen, respectively.

Considering the efficiency of AC/DC inverter ($\eta_{\mathrm{AC/DC}} = 0.95$), the total electricity consumption of an AEC module is expressed as:

$$W_{\mathrm{AEC}} = P_{\mathrm{AEC}} \Delta t / \eta_{\mathrm{AC/DC}} \quad (22)$$

3.4 Hydrogen tank module

The compressed hydrogen tank (HT) for stationary hydrogen storage has been chosen in this work due to its cost-effective performance. The electricity consumption of the HT module (W_{HT}) mainly depends on the compressor, which can be approximated by the following equation:

$$W_{\mathrm{HT}} = \mathrm{w_{compressor}} \max\left(0, \Delta m_{\mathrm{H_2,HT,charge}}\right) \quad (23)$$

where $\mathrm{w_{compressor}}$ denotes the electricity consumption per kilogram of hydrogen compressed and has the value of 4 $\mathrm{kWh \cdot kg^{-1}}$ here [23]. The hydrogen loss of the HT is negligible.

3.5 Proton exchange membrane fuel cell based CHP module

The low-temperature proton exchange membrane fuel cell (PEMFC) typically operates at temperatures below 100°C and has a rapid response and high partial-load efficiency, making it suitable for distributed CHP system close to the households. However, its voltage degradation remains a challenge, seriously affecting its lifetime and performance. Therefore, the lifetime of PEMFC is defined as the time when the cell voltage loss ($\Delta U_{\mathrm{total}}$) reaches 20% of the rated voltage ($\mathrm{U_{cell}^{rated}}$), corresponding to a voltage efficiency (η_{U}) less than $\eta_{\mathrm{U}}^{\mathrm{thre}} = 0.8$. The η_U is defined as follows:

$$\eta_{\mathrm{U}} = 1 - \frac{\Delta U_{\mathrm{total}}}{\mathrm{U_{cell}^{rated}}} \quad (24)$$

The voltage loss is estimated by an operation based degradation model considering four typical operation modes [25], which is expressed as follows:

$$\Delta U_{\mathrm{total}} = \Delta U_{\mathrm{OCV}} + \Delta U_{\mathrm{SSC}} + \Delta U_{\mathrm{VC}} + \Delta U_{\mathrm{CA}} \qquad (\mu V) \tag{25}$$

where ΔU_{OCV}, ΔU_{SSC}, ΔU_{VC} and ΔU_{CA} denote the voltage loss produced from open circuit operation (OCV), start-stop cycles (SSC), voltage cycles (VC) and canlendar aging (CA), respectively. These four types of voltage loss can be assessed by the following equations:

$$\Delta U_{\mathrm{OCV}} = \mathrm{C}_1 \cdot t_{\mathrm{OCV}} \tag{26}$$
$$\Delta U_{\mathrm{SSC}} = \mathrm{C}_2 \cdot n_{\mathrm{SSC}} \tag{27}$$
$$\Delta U_{\mathrm{VC}} = \mathrm{C}_6 \cdot e^{\mathrm{C}_7 \Delta U_{\mathrm{cell}}} \cdot n_{\mathrm{VC}} \tag{28}$$
$$\Delta U_{\mathrm{CA}} = (\mathrm{C}_3 + \mathrm{C}_4 \cdot \mathrm{e}^{\mathrm{C}_5 \cdot U_{\mathrm{cell}}}) \cdot t_{\mathrm{CA}} \tag{29}$$

where t_{OCV} and t_{CA} denote the working time under cell voltage $U_{\mathrm{cell}} > 0.8$ V and constant conditions, respectively. n_{SSC} and n_{VC} denote the number of start-stop cycles and voltage cycles, respectively. The constant parameter C_i uses the validated results from 200 kW Ballard Mark 513 products [25], and has been listed inTable 4.
The cell voltage is calculated by Tafel equation fitted from experimental data (shown in Figure A - 3), which is expressed as:

$$U_{\mathrm{cell}} = 0.779 - 0.137\log(i) \tag{30}$$

where i is the current density, and can be derived from mass flow rate of hydrogen ($\dot{m}_{\mathrm{H_2}}$) consumed using Faraday's law:

$$i = \mathrm{n}\mathcal{F} j_{\mathrm{H_2}} \tag{31}$$

where n corresponds to electron transfer number in electrode reaction and is equal to 2 for hydrogen. $\mathcal{F}$ denotes the Faraday constant which is 96485 $\mathrm{C \cdot mol^{-1}}$. And $j_{\mathrm{H_2}}$ denotes the flux of hydrogen per unit area, and is expressed as:

$$j_{\mathrm{H_2}} = \frac{\dot{m}_{\mathrm{H_2}}}{\mathrm{M_{H_2} A_{stack}}} \tag{32}$$

where $\mathrm{A_{stack}}$ is the total reaction area of stack. For the aforementioned 200 kW PEMFC which has 1320 single cells with reaction area of 300 $\mathrm{cm^2}$, the relationship between hydrogen consumption $\dot{m}_{\mathrm{H_2}}$ and the

percentage (p_{el}) of output electrical power ($P_{\mathrm{el,PEMFC}}$) to rated electrical power ($\mathrm{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated}}$) has been fitted from experimental data (shown in Figure A - 4) and is expressed as:

$$\dot{m}_{\mathrm{H_2}}^{200\ \mathrm{kW}} = 1.476 \times 10^{-6} {p_{\mathrm{el}}}^3 - 3.99 \times 10^{-5} {p_{\mathrm{el}}}^2 + 0.03028 p_{\mathrm{el}} + 0.02724 \quad (33)$$

Given the $\dot{m}_{\mathrm{H_2}}$ under rated electrical power output, we are able to calculate the rated current density (i^{rated}) and then obtain the rated cell voltage ($\mathrm{U}_{\mathrm{cell}}^{\mathrm{rated}}$), in which $i^{\mathrm{rated}} = 1\ \mathrm{A \cdot cm^{-1}}$ and $\mathrm{U}_{\mathrm{cell}}^{\mathrm{rated}} = 0.779\ \mathrm{V}$.

The rated electrical power degradation due to voltage loss can be estimated according to the expression fitted from experimental data (shown in Figure A - 5). The updated $\hat{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated}}$ is expressed as follows:

$$\hat{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated,200\ kW}} = -414.16 \Delta U_{\mathrm{total}} + 200 \quad (34)$$

For convenience, the updating of $\hat{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated}}$ and $\dot{m}_{\mathrm{H_2}}$ have been scaled for calculation under different capacity of PEMFC, as follows:

$$\hat{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated}}(\Delta U_{\mathrm{total}}) = \frac{\hat{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated,200\ kW}}(\Delta U_{\mathrm{total}})}{200} \mathrm{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated}} \quad (35)$$

$$\dot{m}_{\mathrm{H_2}}(p_{\mathrm{el}}) = \frac{\dot{m}_{\mathrm{H_2}}^{200\ \mathrm{kW}}(p_{\mathrm{el}})}{200} \mathrm{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated}} \quad (36)$$

where the percentage of output electrical power $p_{\mathrm{el}} = P_{\mathrm{el,PEMFC}} / \hat{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated}}$.

After the calculation considering cell states of PEMFC, the thermal ($P_{\mathrm{th,PEMFC}}$) and electrical outputs in system level can be derived from partial load efficiency equations fitted from real-world PEMFC systems and thermal utilization ratio [19], as follows:

$$r_{\mathrm{el}} = \frac{\eta_{\mathrm{el}}}{\eta_{\mathrm{el}}^{\mathrm{rated}}} = 1.183 - \frac{0.1756}{p_{\mathrm{el}}} \quad (37)$$

$$r_{\mathrm{th}} = \frac{\eta_{\mathrm{th}}}{\eta_{\mathrm{th}}^{\mathrm{rated}}} = 1.096 - \frac{0.0927}{p_{\mathrm{el}}} \quad (38)$$

$$v_{\mathrm{th}} = \frac{\text{heat extracted}}{\text{heat available}} = \frac{\eta_{\mathrm{th}}^{\mathrm{rated}}}{1 - \eta_{\mathrm{el}}^{\mathrm{rated}}} \quad (39)$$

where r_{el} and r_{th} denote partital load efficiency ratio of electrical and thermal output, respectively.
v_{th} denotes thermal utilization ratio which refers to the ratio of heat extracted to heat available, and is estimated as a constant using rated thermal efficiency ($\eta_{\mathrm{th}}^{\mathrm{rated}} = 0.5$) and rated electrical efficiency ($\eta_{\mathrm{el}}^{\mathrm{rated}} = 0.3$) since there are only electrical and thermal energy produced in electrode reactions. Due to

the approximately linear relationship between electrical efficiency (η_{el}) and cell voltage (U_{cell}) [26], the electrical efficiency considering degradation is expressed as:

$$\eta_{\mathrm{el}} = \eta_{\mathrm{U}}\eta_{\mathrm{el}}^{\mathrm{rated}} \quad (40)$$

Then, the thermal efficiency η_{th} can be calculated as follows:

$$\eta_{\mathrm{th}} = v_{\mathrm{th}}(1-\eta_{\mathrm{el}}) \quad (41)$$

Given the above relationships, it is possible to derive the output electrical power $P_{\mathrm{el,PEMFC}}$ from output thermal power $P_{\mathrm{el,PEMFC}}$ since the PEMFC is designed to meet the heat load as a priority. According to the above equations, the p_{el} and p_{th} (the ratio of $P_{\mathrm{th,PEMFC}}$ to $\mathrm{P}_{\mathrm{th,PEMFC}}^{\mathrm{rated}}$) are related by the following equation:

$$p_{\mathrm{th}} = \frac{P_{\mathrm{th,PEMFC}}}{\mathrm{P}_{\mathrm{th,PEMFC}}^{\mathrm{rated}}} = \frac{\eta_{\mathrm{th}}^{\mathrm{rated}} r_{\mathrm{th}} P_{\mathrm{sys,PEMFC}}}{\eta_{\mathrm{th}}^{\mathrm{rated}} \mathrm{P}_{\mathrm{sys,PEMFC}}^{\mathrm{rated}}} = \frac{\eta_{\mathrm{th}}^{\mathrm{rated}}\left(1.096-\frac{0.0927}{p_{\mathrm{el}}}\right)\frac{P_{\mathrm{el,PEMFC}}}{\eta_{\mathrm{el}}^{\mathrm{rated}}\left(1.183-\frac{0.1756}{p_{\mathrm{el}}}\right)}}{\eta_{\mathrm{th}}^{\mathrm{rated}}\frac{\tilde{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated}}}{\eta_{\mathrm{el}}^{\mathrm{rated}}}} = \frac{\left(1.096-\frac{0.0927}{p_{\mathrm{el}}}\right)}{\left(1.183-\frac{0.1756}{p_{\mathrm{el}}}\right)} p_{\mathrm{el}} \quad (42)$$

then simplified as:

$$1.096p_{\mathrm{el}}^2 - (1.183p_{\mathrm{th}} + 0.0927)p_{\mathrm{el}} + 0.1756p_{\mathrm{th}} = 0 \quad (43)$$

where $p_{\mathrm{th}} \in (0,1)$ and $p_{\mathrm{el}} \in (0,1)$; the values 0 and 1 are excluded as they correspond to shut-down and rated output states, and therefore can be easily resolved.
By solving this quadratic equation, taking p_{th} as a constant while considering the physical mechanism, it yields a solution reflecting relationship between p_{el} and p_{th}, as follows:

$$p_{\mathrm{el}} = \frac{-b+\sqrt{b^2-4ac}}{2a} \quad \text{with } p_{\mathrm{th}} \in (0,0.01629] \cup [0.3771,1) \quad (44)$$

where $a = 1.096$, $b = -(1.183p_{\mathrm{th}} + 0.0927)$, $c = 0.1756p_{\mathrm{th}}$. For convenience, it uses $p_{\mathrm{th}} \in [0.3771,1)$ which yields $p_{\mathrm{el}} \in [0.25,1)$. Therefore, the PEMFC will be turned off when $p_{\mathrm{el}} < 25\%$.
The electricity ($W_{\mathrm{el,PEMFC}}$) produced by PEMFC operating under heat following mode will be transferred to household electricity load (EL) firstly using a DC/AC inverter with efficiency $\eta_{\mathrm{DC/AC}} = 0.95$. If $W_{\mathrm{el,PEMFC}} >$ EL, then the excess electricity will be sold to the grid. It should be noted that, in the PEMFC-CHP module, only the stack of PEMFC is likely to be replaced during the designed lifetime of system ($\mathrm{n}_{\mathrm{sys}}^{\mathrm{life}} = 20$ years).

3.6 Techno-economic parameters

Table 4 summarises the main information of components' parameters. In addition to the technical parameters aforementioned, the economic information includes the initial investment ($cost_i^{ini}$), maintenance cost ($\mathrm{cost}_i^{\mathrm{ma}}$), replacement cost ($\mathrm{cost}_i^{\mathrm{re}}$) and life time ($\mathrm{n}_i^{\mathrm{life}}$) of component i, where the $\mathrm{cost}_i^{\mathrm{re}}$ has considered the recovery value.

Table 4 – Summary of techno-economic parameters.

Module	Parameter	Value
System	$\mathrm{n}_{\mathrm{sys}}^{\mathrm{life}}$	20 year
dToU tariffs	$\mathrm{c}_{\mathrm{e}}^{\mathrm{high}}$	$0.8\ € \cdot \mathrm{kWh}^{-1}$
	$\mathrm{c}_{\mathrm{e}}^{\mathrm{normal}}$	$0.14\ € \cdot \mathrm{kWh}^{-1}$
	$\mathrm{c}_{\mathrm{e}}^{\mathrm{low}}$	$0.0475\ € \cdot \mathrm{kWh}^{-1}$
Heat pump [27]	η_{Lorenz}	0.47
	$\mathrm{T}_{\mathrm{in,sink}}$	313.15 K
	$\mathrm{T}_{\mathrm{out,sink}}$	333.15 K
	$\Delta \mathrm{T}_{\mathrm{air}}$	5 K
	$\mathrm{cost}_{\mathrm{HP}}^{\mathrm{ini}}$	$1400\ € \cdot \mathrm{kW}^{-1}$
	$\mathrm{cost}_{\mathrm{HP}}^{\mathrm{ma}}$	$2\ € \cdot \mathrm{kW}^{-1}$
	$\mathrm{n}_{\mathrm{HP}}^{\mathrm{life}}$	20 year
Hot water tank [23]	$\eta_{\mathrm{loss,\ HWT}}$	0.2% per day
	$\mathrm{cost}_{\mathrm{HWT}}^{\mathrm{ini}}$	$3\ € \cdot \mathrm{kWh}^{-1}$
	$\mathrm{cost}_{\mathrm{HWT}}^{\mathrm{ma}}$	$0.0086\ € \cdot \mathrm{kWh}^{-1}$
	$\mathrm{n}_{\mathrm{HWT}}^{\mathrm{life}}$	20 year
Alkaline electrolytic cell [24]	η_{AEC}	0.665
	$\mathrm{LHV}_{\mathrm{H_2}}$	$241\ \mathrm{kJ} \cdot \mathrm{mol}^{-1}$
	$\mathrm{M}_{\mathrm{H_2}}$	$2\ \mathrm{g} \cdot \mathrm{mol}^{-1}$
	$\eta_{\mathrm{AC/DC}}$	0.95
	$\mathrm{cost}_{\mathrm{AEC}}^{\mathrm{ini}}$	$750\ € \cdot \mathrm{kW}^{-1}$
	$\mathrm{cost}_{\mathrm{AEC}}^{\mathrm{ma}}$	$37.5\ € \cdot \mathrm{kW}^{-1}$
	$\mathrm{n}_{\mathrm{AEC}}^{\mathrm{life}}$	20 year
Hydrogen tank [23]	$\mathrm{w}_{\mathrm{compressor}}$	$4\ \mathrm{kWh} \cdot \mathrm{kg}^{-1}$
	$\mathrm{cost}_{\mathrm{HT}}^{\mathrm{ini}}$	$1900\ € \cdot \mathrm{kg}^{-1}$
	$\mathrm{cost}_{\mathrm{HT}}^{\mathrm{ma}}$	$20\ € \cdot \mathrm{kg}^{-1}$
	$\mathrm{n}_{\mathrm{HT}}^{\mathrm{life}}$	20 year
Proton exchange membrane fuel cell based CHP [25, 28, 29]	C_1	11.4
	C_2	14
	C_3	0.5057
	C_4	0.07866
	C_5	2.965

	C_6	0.0045
	C_7	16.46
	i^{rated}	$1\ A \cdot cm^{-1}$
	U_{cell}^{rated}	0.779 V
	η_{th}^{rated}	0.5
	η_{el}^{rated}	0.3
	η_{sys}^{rated}	0.8
	$\eta_{DC/AC}$	0.95
	$cost_{PEMFC}^{ini}$	$1300\ € \cdot kW^{-1}$
	$cost_{PEMFC}^{ma}$	$17.6\ € \cdot kW^{-1}$
	$cost_{PEMFC}^{re}$	$150\ € \cdot kW^{-1}$
	n_{PEMFC}^{life}	$\eta_U < \eta_U^{thre} = 0.8$

Notes: the currency unit has been converted to Euros (€) according to exchange rate from European Central Bank at 31 December 2021, where 1 $ = 0.88 €, 1 £ = 1.19 €.

4. Modelling of load

4.1 Raw data

All data, including electricity load (EL), heat load (HL), and dToU tariff, are from the Low Carbon London (LCL) project from 2011 to 2014, where the DR trials [30, 31] provide ELs from treatment households using dToU tariffs and control households using a fixed tariff, in a 30-minute time resolution, and the heat pump (HP) trial [32] provides HLs from several households not taken part in DR trial, in a 15-minute time resolution. The following sub-sections describe their relevant information and the processing in detail.

4.1.1 Time-of-Use tariffs

During the LCL project period, 2013 was selected for the DR trial, with 153 days set as dToU pricing events for network constraints and RE power supply following purposes. The 5566 households were divided into a treatment group (20%) and a control group. In the treatment group, tariff prices were given one day in advance via the Smart Meter In-Home Display (IHD) or text message. Customers receive price signals for high (67.20 pence/kWh), low (3.99 pence/kWh) or normal (11.76 pence/kWh) prices and the hours at which these signals apply. A fixed rate of 14.228 pence/kWh was used for the control group. Figure A - 1 and Figure A - 2 describe the temporal distribution of dToU tariffs. In general, high tariffs are concentrated in the peak hours (17-23 pm), while low tariffs are mainly distributed in the off-peak hours. In the following optimization, the dToU tariffs representing demand response are taken as fixed variables as they are prespecified 24 hours ahead.

4.1.2 Load profiles

The total electricity consumption of 242 treatment households with complete records for 2013 was selected to stand for the load profile of a district energy system for community scale, which corresponds to an average electricity consumption of 90.65 kWh per hour.

The heat load profiles of 6 households beyond DR trial have been collected with records of operational data from heating system mainly driven by heat pump, during periods from Dec. 2011 to Mar. 2014. The heating system is designed to meet the demand of space heating (SH) and domestic hot water (DHW), and is operated mainly in intermittent or continuous modes.

The operational data includes outlet (T_{out}, °C) and inlet (T_{in}, °C) water temperature of the heat pump, flow rate ($\dot{V}$, $m^3 \cdot h^{-1}$), outside and inside air temperature of the property. Based on these parameters, we can calculate household heat load as follow:

$$HL^t = \dot{V}\rho c_p(T_{out}^t - T_{in}^t) \quad (45)$$

where ρ and c_p denote the density and heat capacity of water, respectively.

For enriching the representativeness of load profiles, the heat loads in 2012 and 2014 available in 5 households have been readjusted separately considering the matching of holidays and days of week compared to that of 2013, to form 5 load profiles of a complete year. Together with 6 load profiles in 2013, it provides 11 heat load profiles in total, which has been then scaled up to 242 households to match the community scale. The mean hourly consumption of total heat load is 199.69 kWh, which is 2.2 times of electricity load. Figure 3 shows the mean heat and electricity loads throughout the day.

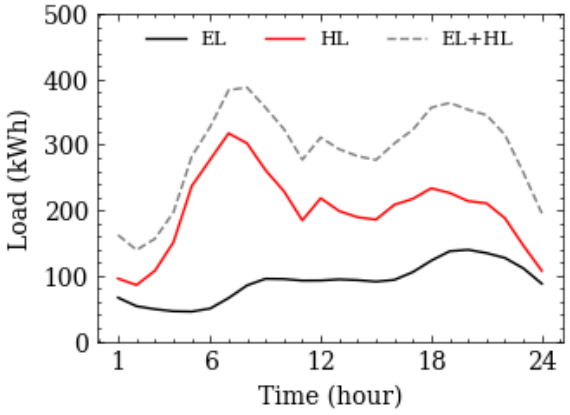


Figure 3 – Mean demand loads over a 24 hour period. (The black curve indicates the Electricity load (EL), the red curve indicates the Heat load (HL), and the dashed curve indicates the combined load (EL+HL)).

The time order of total heat load has been also readjusted to represent a complete heating season for the purposes of simulation, in which the heat and electricity loads in range from 2013.1.1 to 2013.7.30 are moved to behind 2013.12.31. In what follows, the optimization will be based on these readjusted load

profiles. While, the raw time order load profiles will be used for building the load prediction model and testing the generalization performance of model predictive control and reinforcement learning models.

4.2 Load prediction

Accurate load prediction can provide the operation strategies with essential information for meeting specific targets, such as cost reduction. However, due to the complex and stochastic energy consumption behaviour from people in household, there exists unavoided uncertainty when electricity and heat load is forecasted [33, 34]. Moreover, the uncertainty increases with the length of the prediction horizon, as well as the reduction of consumption scale [35]. As a result, a need has arisen for probabilistic prediction model with high accuracy and measurement of uncertainty [36].

In response to this, this paper combines quantile regression [37] with long-short term memory neural network (LSTM) [38] to provide accurate interval prediction of load; we will refer to this approach as the QLSTM model. The LSTM improved from recurrent neural network is able to learn the dependencies between output and time series inputs, which makes it one of the most effective machine learning models for load prediction [39]. Based on the LSTM framework, the quantile regression offers a loss function to forecast the given quantile value of load, without prespecifying the distribution of load. A detailed description of QLSTM has been summarized in Appendix C.1.

In this work, the 5% and 95% quantiles were chosen to achieve a 90% interval prediction of load. The 50% quantile prediction is taken as the point prediction. Different prediction horizon lengths were considered, with values ranging in the set {0.5,1,4,8,12,16,20,24} (hours), in order to explore their impacts on operation optimization effects. Table 5 lists the main information required for training and testing of QLSTM. In the evaluation stage, the mean pinball (PB) and mean Winkler score (WS) for interval prediction, and mean absolute percentage error (MAPE) and root mean squared error (RMSE) for point prediction have been selected and summaried in Table A - 1, where lower values mean better performance. All the prediction models were built using the machine learning package tensorflow (v2.2.0), running on a machine with a NVIDIA K80 GPU.

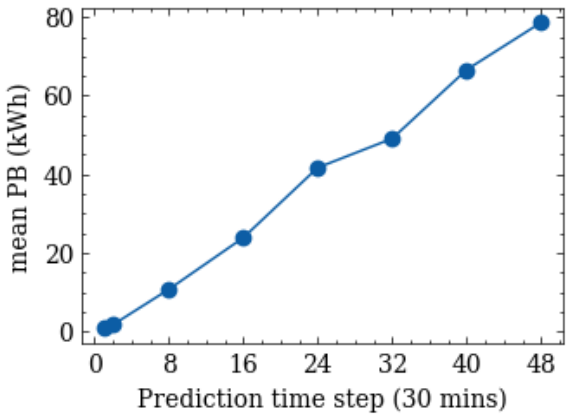
Table 5 – Main information of QLSTM training and test.

Load type	Electricity load	Heat load
Dataset division	Training/valid/test set: 8/2/2 months	
Time resolution	30 mins	15 mins
Features and preprocessing	Load: N(0,1) standardization • Total electricity load • dToU tariffs Weather: N(0,1) standardization • Temperature • Sunrise time Calendar features: OneHot encoding • If holiday	Load: N(0,1) standardization • Total heat load 11 profiles: N(0,1) standardization • Heat load • Outside air Temperature • Inside air Temperature Calendar features: OneHot encoding • If holiday

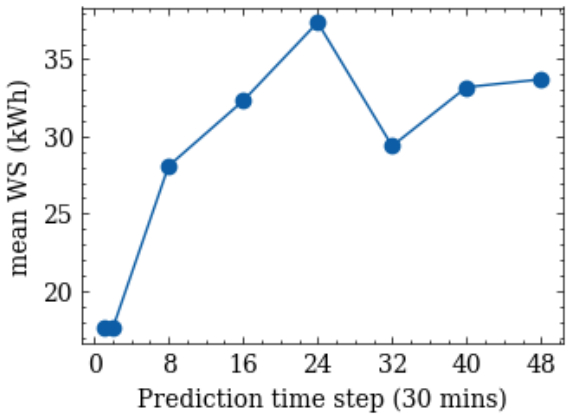
	• Day of a week • Half-hour of a day	• Day of a week • Quarter-hour of a day
Regression variable	$Y_{q,t} \in \mathrm{R}^{3\times t}$: q-quantile of loads at each time step in future t steps (T hours) where $q = \{5\%, 50\%, 95\%\}$, $\mathrm{T} = \{0.5,1,4,8,12,16,20,24\}$ hours. Electricity load: $\mathrm{t_{EL}} = 2\mathrm{T}$; Heat load: $\mathrm{t_{HL}} = 4\mathrm{T}$.	
Hyperparameter optimization	Bayesian optimization: Layers: {1,2,3} Units: {64, 128, 256} Training time steps: 1-7 past days $\mathrm{t_{train}} = 48 \times \{1,2,3,4,5,6,7\}$	Grid search: Layers: 1 Units: 128 Training time steps: 0.5-3 past days $\mathrm{t_{train}} = 96 \times \{0.5,1,1.5,2,2.5,3\}$
Training setting	Nadam optimizer; Earlystopping with Δ=1e-6 for valid loss over 10 steps	
Loss value	Mean value of 5%, 50% and 95% quantile loss	
Evaluation criteria	Interval prediction: mean PB, mean WS; point prediction: MAPE, RMSE	

Notes: For the heat load prediction model, the values of layer and unit share the optimization results of electricity load prediction, and the training time steps have been limited to 0.5-3 past days due to the limitation of computational resources.

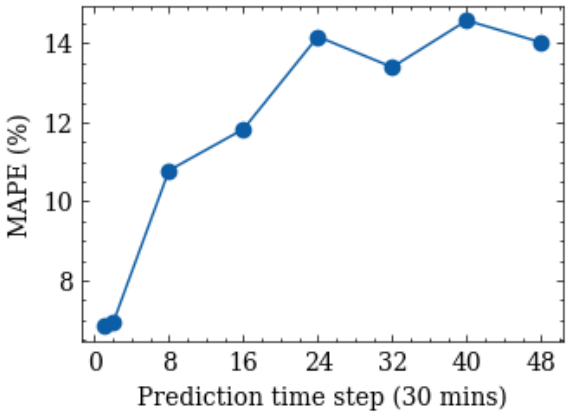
Figure 4 and Figure 5 reveal the evaluation results of trained QLSTM models using optimal hyperparameters (shown in Figure A - 7) on test dataset of electricity and heat load, respectively. The detailed computation data can be found in Table A - 2 and Table A - 3. It can be generally concluded that the prediction error increases with the rise of prediction horizon, both in point and interval prediction. While the growth rate of the prediction error is reduced with the increase in the prediction horizon for point prediction and is almost constant for interval prediction.



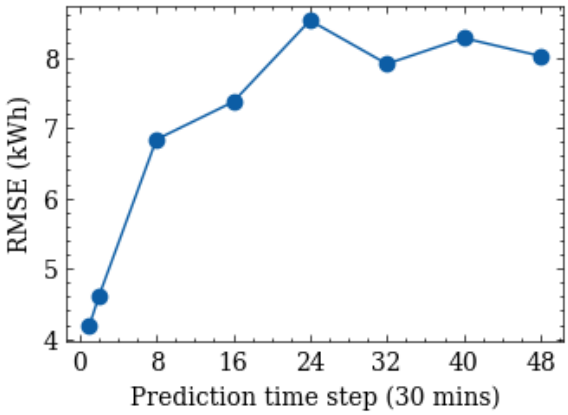
(a) Prediction error (interval prediction)



(b) Interval size (interval prediction)

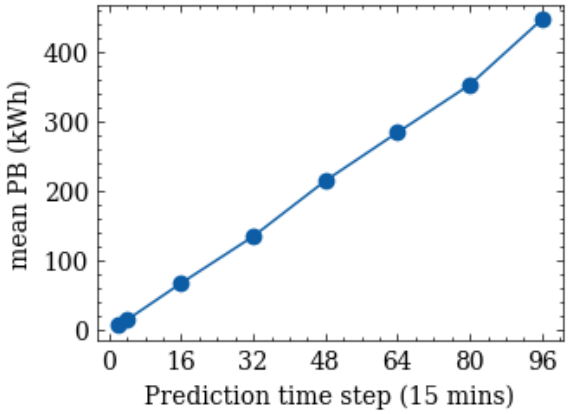


(c) Relative error (point prediction)

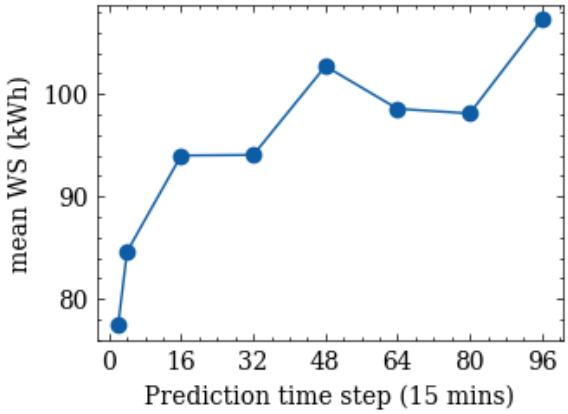


(d) Absolute error (point prediction)

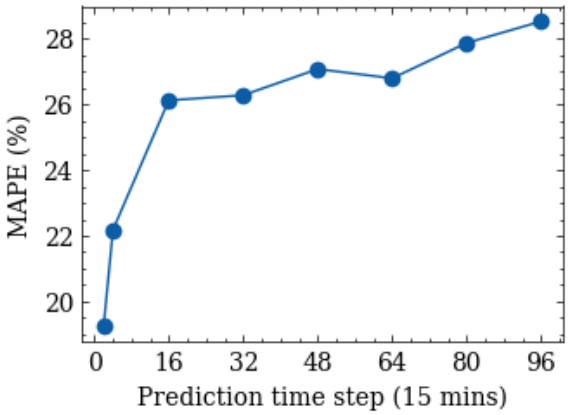
Figure 4 – Evaluation of trained QLSTM models on test dataset for electricity load.



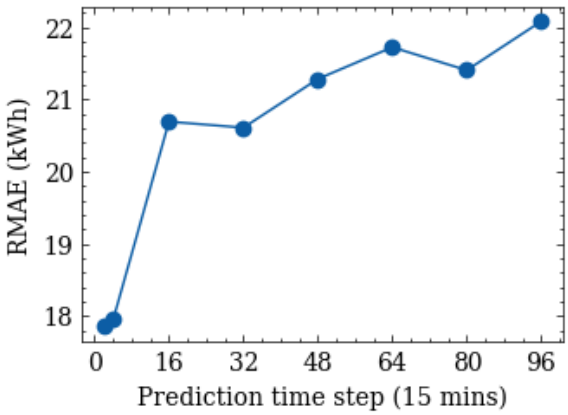
(a) Prediction error (interval prediction)



(b) Interval size (interval prediction)



(c) Relative error (point prediction)



(d) Absolute error (point prediction)

Figure 5 – Evaluation of trained QLSTM models on test dataset for heat load.

5. Optimization methods

5.1 Capacity optimization

Given the defined capacity optimization problems with the form of mixed integer nonlinear programming (MINLP), the particle swarm optimization (PSO) in swarm intelligence and Bayesian optimization (BO) in data-driven methods have been used to find the optimal solutions. These two kinds of methods are both free in the form of optimization problem and variable but follow different optimization strategies, which makes them outperform traditional methods for MINLP problems. The detailed descriptions of PSO and BO have been illustrated in Appendix C.2 and C.3, respectively. A performance comparison between PSO and BO is beneficial for selecting the suitable algorithm when faced with more complex energy system designs in the future.

The PSO is carried out based on our own developed python package gensbo [40], while the BO is based on the package hyperopt [41]. Table 6 lists the key parameters of the two algorithms. For the capacity optimization of S_{B0}, S_{B1} and S_{C1}, 10 independent calculations have been run on a desktop machine with an Intel i7-6700 CPU @ 3.40 GHz and 16 GB of RAM.

Table 6 – Key parameters of PSO and BO

Method	PSO	BO
Parameter	Swarm size: 50; max steps:100	Max steps: 5000

Notes: other parameters use the default setting from the packages.

5.2 Operation optimization

For optimizing the operation strategies, model predictive control based on mathematical programming and reinforcement learning based on data-driven models have been selected. The basic idea behind the model predictive control is to convert control problems into optimization problems, where the optimization problem is solved online in a finite time domain with the current system state as the initial state, and the solution is the control action at that time [42]. It relies on rolling prediction which provides future information for calculation. In S_B and S_C scenarios, the modelling of MPC can be illustrated as a process to construct an optimization problem based on initial information and then to find the optimal output of source components at the next time step, on a rolling basis, under specific objectives.

Reinforcement learning is a balancing game of exploration and exploitation; the former allows an agent to fully explore the possibilities in the environment and thus has the opportunity to find the optimal

solution; the latter uses the learned experience to guide the agent to make more rational choices [43]. The model-free RL algorithm named soft actor-critic [44] is chosen to optimize operation due to its superior learning ability and good stability in training for continuous space tasks [45]. Based on records from the interaction between agent and environment, it learns a policy ($\pi(s)$), which guides the action (a) of agent under current state (s), and value functions ($Q(s,a)$) estimating rewards from the environment. The detailed description of SAC has been provided in Appendix C.4. In response to scenarios S_B and S_C, the working process of RL can be described as follows: in the environment (integrated energy system), the agent (source components), faced with the current state s^t (initial information), makes an action a^t (output) in order to expect the maximum reward $Q(s^t, a^t)$ through the experience $\pi(s^t)$ obtained through previous learning (off-line training), and then it is found that the true feedback is $R(s^t, a^t)$.

Given the initial conditions at time step t and information within future time steps $T = 4 \times \{0.5,1,4,8,12,16,20,24\}$ (listed in Table 7), MPC finds optimal control actions by constructing a online optimization problem in a finite time domain, while the RL provides solutions from a policy model trained off-line. For both methods, the optimal actions at the following time step, $t + 1$, will be used to control the source components. Besides, the minimum time resolution between electricity and heat load has been used to represent the length of time step, which means $\Delta t = 15$ mins. Therefore, the electricity load value corresponding to each time step will be taken as half of the original electricity load value.

Table 7 – Initial information used for MPC and RL.

Initial information		MPC		RL	
		S_{B2}	S_{C2}	S_{B3}	S_{C3}
Time step t	Q_{HL}^t (kWh)			√	√
	$LHSP_{HL}^t$ (kW)			√	√
	$T_{outside}^t$ (K)			√	√
	COP_{HP}^t			√	
	c_e^t ($€ \cdot kWh^{-1}$)			√	√
	SoC_{HWT}^t	√	√	√	√
	$c_{h,HWT,discharge}^t$ ($€ \cdot kWh^{-1}$)	√	√	√	√
	SoC_{HT}^t		√		√
	$c_{H_2,HT,discharge}^t$ ($€ \cdot kg^{-1}$)		√		√
	$\Delta U_{total,PEMFC}^t$ (V)		√		√
	$U_{cell,PEMFC}^t$ (V)		√		√
	$p_{el,PEMFC}^t$ (%)				√
Future time step T. $j = \{1, \dots, T\}$, $T = 4 \times \{0.5,1,4,8,12,16,20,24\}$	$\widehat{W}_{EL}^{t+j}$ (kWh)	√	√	√	√
	$\widehat{Q}_{HL}^{t+j}$ (kWh)	√	√	√	√
	c_e^{t+j} ($€ \cdot kWh^{-1}$)	√	√	√	√
	$T_{outside}^{t+j}$ (K)			√	√

$\mathrm{COP}_{\mathrm{HP}}^{t+j}$	√	√

Notes: the outside temperatures ($\mathrm{T}_{\mathrm{outside}}^{t+j}$) within future T time steps use true records from Heathrow Airport in London. And for RL, all the variables listed are transferred into state space by preprocessing with [0,1] interval scaling.

The discharge price of heat from HWT ($\mathrm{c}_{\mathrm{HWT,discharge}}^{t}$) and hydrogen from HT ($\mathrm{c}_{\mathrm{HT,discharge}}^{t}$) are represented by average charge cost over the past 96 time steps (i.e. the past 24 hours) if charge action happens, which can be calculated as:

$$\mathrm{c}_{\mathrm{h,HWT,discharge}}^{t} = \frac{\sum_{j=1}^{96} \mathrm{cost}_{\mathrm{HWT,charge}}^{t-j}}{\sum_{j=1}^{96} \mathrm{Q}_{\mathrm{HWT,charge}}^{t-j} (1-\eta_{\mathrm{loss,\ HWT}})^{j}} \quad (46)$$

$$\mathrm{c}_{\mathrm{H_2,HT,discharge}}^{t} = \frac{\sum_{j=1}^{96} \mathrm{cost}_{\mathrm{HT,charge}}^{t-j}}{\sum_{j=1}^{96} \mathrm{m}_{\mathrm{HT,charge}}^{t-j}} \quad (47)$$

In addition, there are three different kinds of predictive loads, in both electricity and heat load, which includes true load, point prediction results, and interval prediction results. The detailed information of predictive loads has been summarized in Table 8.

Table 8 – Information of predictive load for operation optimizaztion.

Prediction model	True load	Point prediction	Interval prediction
Predictive quantile q	-	{0.5}	{0.05, 0.5, 0.95}
Predictive time steps	$\{1, \ldots, T\}$, where $T = 4 \times \{0.5,1,4,8,12,16,20,24\}$		
Size	$1 \times T$	$1 \times T$	$3 \times T$

For solving MPC, the Gurobi solver is used to solve the built nonlinear programming (NLP) problems on a desktop machine with an Intel i7-6700 CPU @ 3.40 GHz and 16 GB of RAM. The maximum running time for each time step in S_{C2} has been limited to 60 s.

The RL is trained 5 times independently in adjusted order load profiles (corresponding to a complete heating season), using the python package tianshou [46] on a desktop machine with an Intel i7-6700 CPU @ 3.40 GHz and 16 GB of RAM. Table 9 lists the main settings for SAC training. Figure A - 6 shows two samples of training process. The training process will automatically save the policy model with the best value of the reward function in the evaluation session to achieve a better balance between fitting and generalization.

Table 9 – Main settings of SAC.

Scenario	S_{B3}	S_{C3}

Policy model; Value function model	Multilayer perceptron (MLP)； Layers:2; Units: 512 for point predictive load and 1024 for interval predictive loads	
Max epochs for training	75	250

Notes: other parameters use the default setting from tianshou package.

6. Results and discussion

With the given settings of the PSO and BO for capacity optimization (S_{B0}, S_{B1}, S_{C1}), the MPC (S_{B2}, S_{C2}) and RL (S_{B3}, S_{C3}) for operation optimization, the adjusted time order load profiles with a complete heating season are provided as the basis for system optimization, while the raw time order load profiles are used to evaluate the generalization ability of MPC and RL. In order to better compare the performance of optimization algorithms, the PSO, BO, MPC, RL are implemented on a desktop computer with i7-6700 CPU (3.40 GHz) and RAM 16 GB, in which parallel computing was not employed.

6.1 Results of optimization

Given the load profiles and system structure, the capacity optimization follows a passive heat load following strategy to provide optimal solutions for operation optimization. Furthermore, combined with dToU tariffs and load prediction, active load matching is carried out to reduce overall cost and provide flexibility.

6.1.1 Capacity optimization

Table 10 lists the detailed results of the capacity optimization, while Table 11 reveals the comparison of best results between S_{B0} (HP), S_{B1} (HP+HWT), and S_{C1} (AEC+HT+PEMFC+HWT). In the optimization results aimed at economics, the S_{B0} system without thermal storage requires a 200 kW HP to meet the heating load, but the rated power of the HP can be reduced to 150 kW after introducing HWT. In addition, under the passive heat-led strategy, the optimization results show that the threshold ($SoC_{\mathrm{HWT}}^{\mathrm{thre}}$) used to control the charging of HWT has no significant difference between the heating and non-heating seasons, and both require HWT to be at a high SoC level to cope with peak loads. In the hydrogen-based scenario S_{C1}, the optimal results obtained using PSO and BO show some differences in $SoC_{\mathrm{HWT}}^{\mathrm{thre}}$ between the heating and non-heating seasons, but both are still at a relatively high level. Considering that the optimization results obtained from the two intelligent optimization methods have certain randomness, the best result for the complex scenario S_{C1} may only be a local optimal solution rather than a global optimal solution. In the following, the analysis will be based on the solutions with minimum COE.

Generally, S_{B1} has the lowest COE, and S_{C1} is over two times of S_B. Besides, the operation cost accounts for over 90% of COE in S_B, which drops to 67.6% in S_C due to its costly components. The higher share of operation cost in S_{B1} is likely to offer greater potential for operation optimization, which highlights the importance of storage since the cost-effective storage components are able to extend the optimization

space (i.e. the flexibility).

Table 10 – Best results of capacity optimization (10 runs).

	S_{B0}	S_{B1}		S_{C1}	
	PSO & BO	PSO	BO	PSO	BO
COE(€/kWh)	0.094016	0.091863	0.091857	0.22137	0.22100
LHSP	0.0076	0.0075	0.0072	0.0056	0.0096
$P_{\mathrm{HP}}^{\mathrm{rated}}$ (kW)	200	150	150		
$Q_{\mathrm{HWT}}^{\mathrm{rated}}$ (kWh)		1210	1340	15480	8430
$SoC_{\mathrm{HWT,DHW}}^{\mathrm{thre}}$		0.95	0.95	0.85	0.9
$SoC_{\mathrm{HWT,SH}}^{\mathrm{thre}}$		1	1	0.85	0.7
$P_{\mathrm{el,PEMFC}}^{\mathrm{rated}}$ (kW)				670	670
$P_{\mathrm{AEL}}^{\mathrm{rated}}$ (kW)				480	480
$m_{\mathrm{HT}}^{\mathrm{rated}}$ (kg)				70	80
$SoC_{\mathrm{HT,DHW}}^{\mathrm{thre}}$,				0	0
$SoC_{\mathrm{HT,SH}}^{\mathrm{thre}}$				0	0
Number of PEMFC replacement				10	10

Notes: In order to ensure the accuracy of the calculation and to facilitate comparison, the calculation results of COE are taken as 5 valid digits and the rest of the calculation results are taken as 2 valid digits (similarly hereinafter).

Table 11 – Comparison between best results of capacity optimization in S_{B0}, S_{B1} and S_{C1}.

	S_{B0}	S_{B1} (BO)	S_{C1} (BO)	$(S_{B1}-S_{B0})/S_{B0}$	$(S_{C1}-S_{B0})/S_{B0}$	$(S_{C1}-S_{B1})/S_{B1}$
COE (€/kWh)	0.094016	0.091857	0.22100	-2.29 %	+135.08 %	+140.59 %
$\mathrm{COE}_{\mathrm{operation}}$ (€/kWh)	0.084977	0.084947	0.14934	-0.00%	+75.74%	+75.81%
$\mathrm{COE}_{\mathrm{operation}}/\mathrm{COE}$	0.90	0.92	0.68	2.32%	-25.20%	-26.90%

6.1.2 Operation optimization

The optimal results of operation in S_B and S_C have been summarized in Table 12 and Table 13, respectively. Detailed computation results under different prediction horizons can be found in Appendix A (form Figure A - 8 to Figure A - 11, and from Table A - 4 to Table A - 20). It's noted that the satisfaction of LSHP constraint and good generalization on raw time order load profiles are the necessary conditions. Therefore, the MPC produces the outperformed results in S_B (S_{B2}) by using interval prediction of load within future 16 hours. While the RL provides the best optimal solutions in S_C (S_{C3}) under the input of point prediction of load within future 20 hours. With the help of operation optimization, both S_B and S_C realize obvious reduction of COE, with respect to -5.92% and -8.31% as compared to S_{B0} and S_{C1}, respectively, in which the number of PEMFC including replacement of stack has been effectively decreased from 10 to 6,

corresponding to a reduction of 40%.

Table 12 – Comparison between optimal results of S_B.

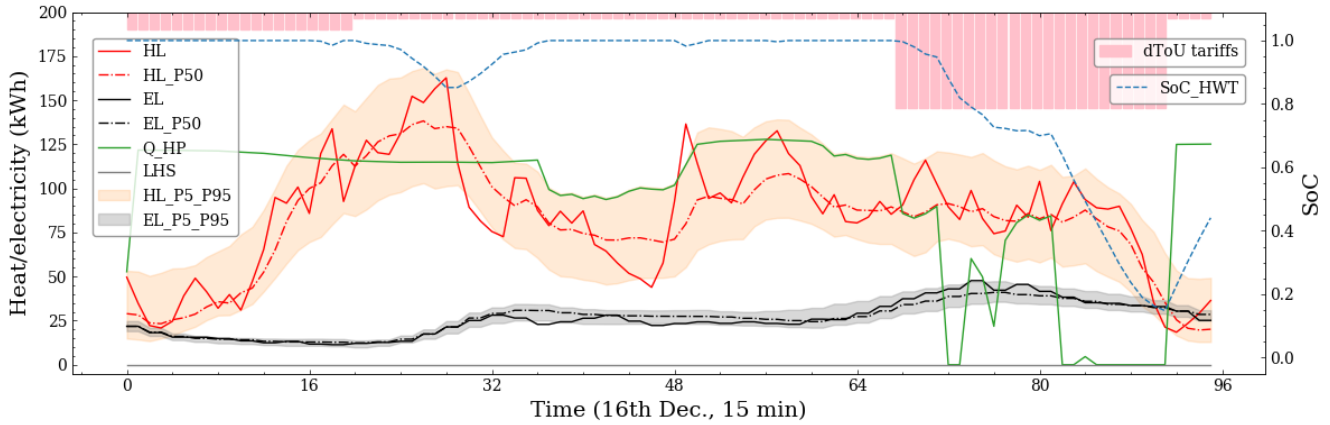
		S_{B0}	S_{B1}	S_{B2}	S_{B3}	$(S_{B1}-S_{B0})/S_{B0}$	$(S_{B2}-S_{B0})/S_{B0}$	$(S_{B3}-S_{B0})/S_{B0}$
LHSP		0.0076	0.0072	0.0078	0.0089			
COE (€/kWh)		0.094016	0.091857	0.088442	0.088020 (Best of 5 runs)	-2.29%	-5.92%	-6.37%
$COE_{operation}$ (€/kWh)		0.084977	0.084947	0.081457	0.081025	-0.04%	-4.14%	-4.65%
$COE_{operation}/COE$		0.90	0.92	0.92	0.92	+2.31%	+1.89%	+1.84%
Prediction horizon (hour)				16 (interval prediction)	20 (point prediction)			
Test of generalization	LHSP			0.0078	0.0075			
	COE (€/kWh)			0.088447	Unsatisfied (0.098773)			

Table 13 – Comparison between optimal results of S_C.

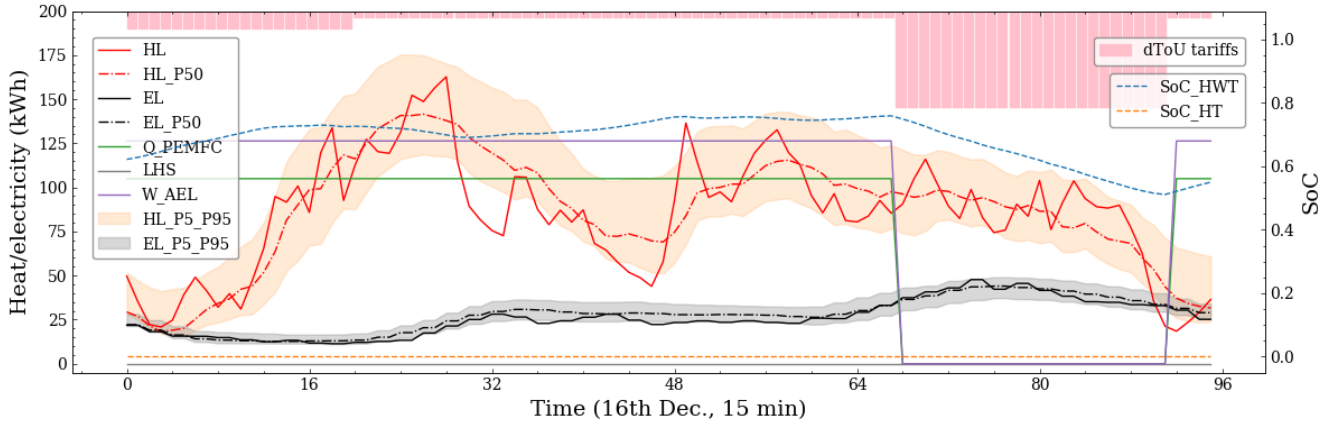
		S_{B0}	S_{C1}	S_{C2}	S_{C3}	$(S_{C1}-S_{B0})/S_{B0}$	$(S_{C3}-S_{B0})/S_{B0}$	$(S_{C3}-S_{C1})/S_{C1}$
LHSP		0.0076	0.0096	Unsatisfied (0.011)	0.0098			
COE (€/kWh)		0.094016	0.22100	(0.21675)	0.20264 (Best of 5 runs)	135.08%	115.55%	-8.31%
$COE_{operation}$ (€/kWh)		0.084977	0.14934	-	0.13014	75.75%	53.15%	-12.86%
$COE_{operation}/COE$		0.90	0.68	-	0.64	-25.24%	-28.95%	-5.03%
Number of PEMFC			10	-	6			-40%
Prediction horizon (hour)				-	20 (point prediction)			
Test of generalization	LHSP			-	0.0084			
	COE (€/kWh)			-	0.20263			
	Number of PEMFC				6			

Figure 6 presents the system operation diagrams of the S_{B2} and S_{C3} for a typical day in heating season with high load, under the dToU tariffs, respectively. The pink (heat load) and gray (electricity load) bands

displayed in the figure represent the rolling load prediction range given by QLSTM for a specified prediction horizon, indicating that the 5 % and 95 % percentile prediction intervals effectively cover the majority of actual load. Given the information of predicted load, dToU tariffs and device states, the optimization make source devices (HP, AEC, and PEMFC) work less during high tariff hours and continuously during low tariff hours in order to meet the energy supply. This trend is evident in the RL-optimized Sc3 scenario, but is weaker in the MPC-optimized Sb2 scenario. This may be due to the fact that the stochastic optimization of MPC needs to consider higher values of 95 % quantile load prediction scenarios, making it give a more conservative control strategy. In addition, the PEMFC does not work at high tariffs but works at low tariffs, which does not meet the objective of maximizing economic benefits, probably also due to the consideration of coping with load uncertainty to keep the HWT of this system at a relatively high and stable SoC. Compared to the operational optimization without considering load uncertainty, the optimization strategy tends to give relatively conservative control actions when load uncertainty is considered, rather than a single pursuit of cost minimization, which can be verified in the later analysis of the impact of load uncertainty on the optimization effect. Nevertheless, the overall results show that both schemes can effectively use dynamic tariffs to reduce system operating costs.



(a) S_{B2} using MPC given load from interval prediction with prediction horizon of 16 hours.



(b) S_{C3} using RL given load from point prediction with prediction horizon of 20 hours.

Figure 6 – Visualization of optimal system operation of S_{B2} and S_{C3} in a typical winter day with high heat and

electricity load under dToU tariffs.

6.2 Results of economical effects

Based on the optimization of capacity and operation, the optimal results of S_B and S_C have been summarized in Table 14. Compared to grid-based HP system without heat storage (S_{B0}), the extension of HWT and the optimization of capacity (S_{B1}) reduced 2.29% of the overall COE. Optimization of operation (S_{B2}) achieves a further 5.92% reduction. Considering the high share of operation cost ($\mathrm{COE_{operation}}$), the comparison of $\mathrm{COE_{operation}}$ between S_{B0} and S_{B2} demonstrates a decrease of 4.14% due to the introduction of operation optimization.

In contrast to the outperformed short term scenario S_{B2}, the long term scenario S_{C3} costs around 1.3 times higher in overall COE. In detail, $\mathrm{COE_{operation}}$ of S_{C3} is about 0.6 times higher than that of S_{B2}. From the perspective of energy conversion efficiency, this suggests that the overall efficiency of AEC-PEMFC needs to be improved by 0.6 times to be competitive with a HP system. As an example, the energy conversion efficiencies of the AEC and PEMFC in this work are 0.665 and 0.8, respectively, with an overall efficiency of 0.532, which would be 0.85 after multiplying it by 1.6 times. This approximately corresponds to an efficiency of 0.9 and 0.95 for the AEC and PEMFC, respectively. In terms of equipment costs $\mathrm{COE_{equipment}}$ (including all other costs except $\mathrm{COE_{operation}}$), S_{C3} would probably require a reduction to 1/10 of its current level to be comparable to S_{B2}. According to Figure 7, the PEMFC (68%) and AEC (24%) account for over 90% of $\mathrm{COE_{equipment}}$, which indicates that the high cost of PEMFC-CHP and AEC is one of the main reasons that hinders the application and rollout of S_C.

Table 14 – Comparison of best results between S_B and S_C.

	S_{B0}	S_{B2}	S_{C3}	$(S_{B2}-S_{B0})/S_{B0}$	$(S_{C3}-S_{B0})/S_{B0}$	$(S_{B3}-S_{B2})/S_{B2}$
COE (€/kWh)	0.094016	0.088442	0.202640	-5.92%	115.55%	129.12%
$\mathrm{COE_{operation}}$ (€/kWh)	0.084977	0.081457	0.130144	-4.14%	53.15%	59.77%
$\mathrm{COE_{operation}/COE}$	0.90	0.92	0.64	1.89%	-28.95%	-30.29%
$\mathrm{COE_{equipment}}$ (€/kWh)	0.0090330	0.0069851	0.072496	-22.67%	702.57%	937.88%

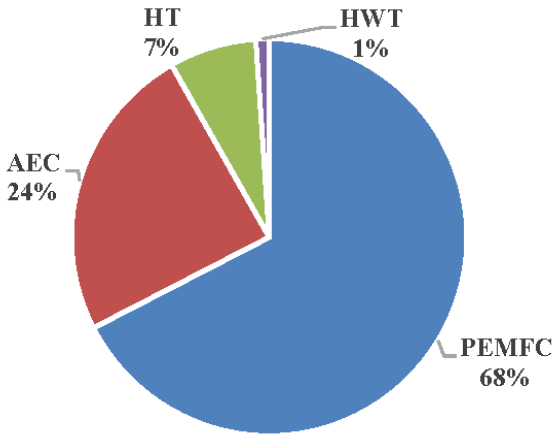


Figure 7 – Breakdown of equipment cost (including initial inverst, maintenance and replacment) for components in S_{C3}.

6.3 Results of optimal flexibility

The high and low tariffs, as important signals to guide load leveling and renewable power consumption, are the core parameters for the operation optimization carried out in this study. Table 15 accounts the electricity consumption of the energy system under high and low dToU tariffs. Compared with the electricity load of capacity optimization (C) without operation optimization (O), the operation optimization in the S_B decreased about 44.98% during high tariff periods and increased about 140.91% during low tariff periods. The regulation effect in the S_C was more significant, and its operational optimization reduced the high-priced electricity load by 97.40% and increased the low-priced electricity consumption by 153.97%. Obviously, great improvements in operation optimization results are observed compared to capacity optimization results, with the S_C showing higher regulation ability, which demonstrates that the operation optimization provide effective flexibility under the dToU tariffs designed for network constraint and RE power supply following.

Table 15 – Distribution of energy system electricity consumption under high and low tariffs.

Tariff type	S_B		S_C	
	High	Low	High	low
C: capacity optimization (MWh)	21.93	36.69	62.84	107.17
O: operation optimization (MWh)	12.06	88.39	1.63	272.18
Relative change: (C-O)/C	-44.98%	+140.91%	-97.40%	+153.97%

The optimization also leads to an obvious reshape of electricity load curve. To visually demonstrate this change, comparison of mean electricity load curve between raw data and optimal solutions of S_B and S_C has been presented in Figure 8, where the raw data of non-ToU represents the mean load in 2013 of control households under a fixed tariff. The results reveal a small reduction of electricity consumption in treatment

households due to the introduction of dToU tariffs in contrast to a fixed tariff, while the peak still locates in evening periods. However, when delivering the heat load by grid-based system, the shape and magnitude of the electricity load will be greatly affected in the meantime. Specifically, the S_{B2} using heat pump leads to an increase of electricity consumption and a double-peak load curve by increasing the morning peak. The S_{C3} based on hydrogen greatly increases the consumption and moves the peak from evening to early morning periods. The reasons for the changes can be mainly attributed to the high COP of HP (>1) and low COP of hydrogen based system (<1), as well as the introduction of storage and dToU tariffs oriented operation optimization, which can be clearly observed by comparing the mean electricity load curve by percentage of different solutions within the same scenario, as shown in Figure 9.

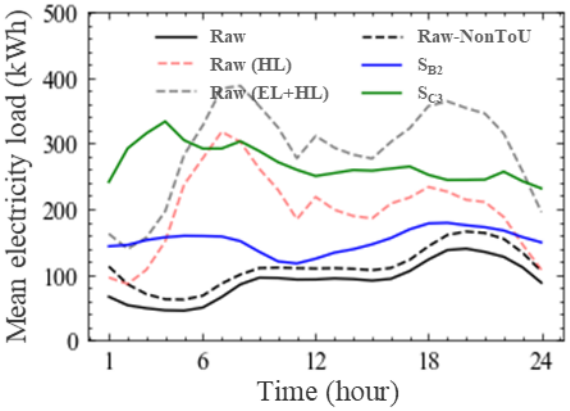
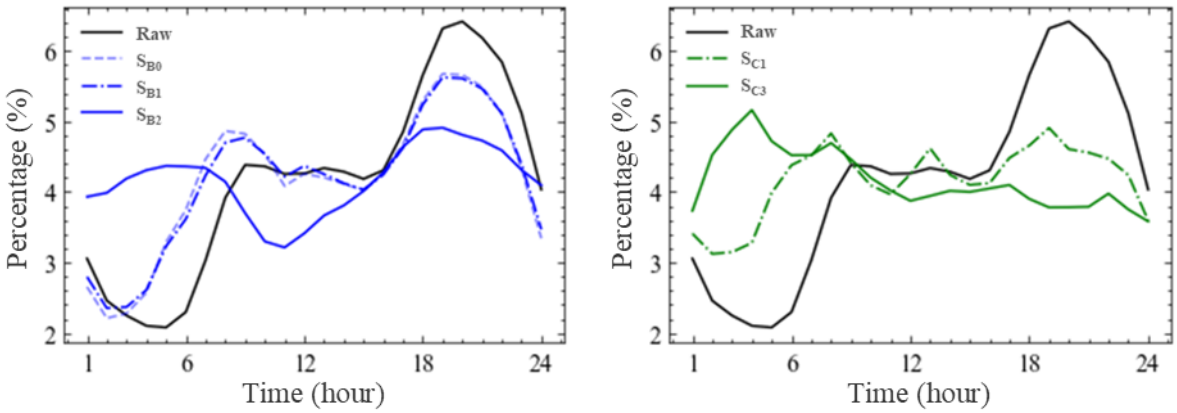


Figure 8 – Comparison of mean load shape between different scenarios.



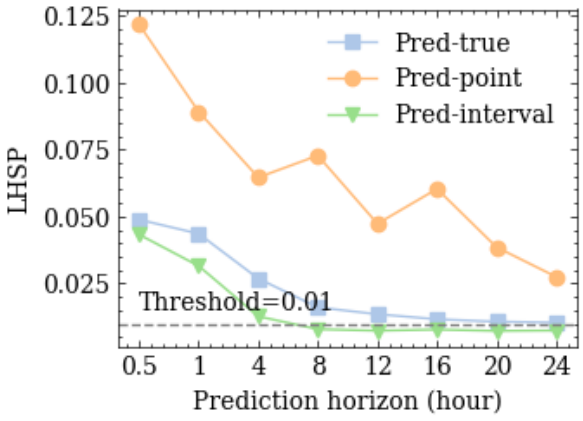
(a) S_B (b) S_C

Figure 9 – Comparison of mean electricity load shape by percentage between different solutions in the same scenario.

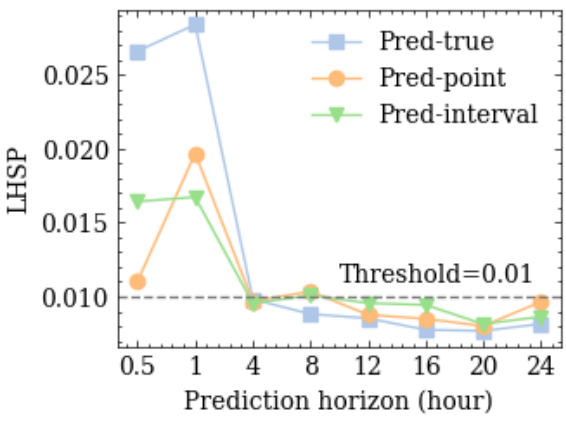
6.4 Results of impacts from load uncertainty on optimization effect

As a major source of uncertainty, load information has a great influence on the effects of operation optimization. The LHSP and COE results for operation optimization using true load, point prediction load, and interval prediction load for different prediction horizons are shown in Figure 10 and Figure 11. Firstly, the results show that the optimization effect using true load is generally higher than the optimization effect using predicted load containing errors, meaning that the load accuracy has a significant impact on the optimization effect. Secondly, in terms of prediction horizon of load, it is observed that the LHSP decreases as the length of prediction horizon increases, while the COE decreases and then increases with increasing prediction horizon. With the length increases in the prediction horizon, the increased load information together with rising prediction error make the benefit from load prediction brought to optimization suffer from increase to decrease. In addition, considering the LHSP constraint and optimization objective under the long-term prediction horizon makes the energy reserve of the system at the next moment more adequate than when only the short-term prediction horizon case is considered, which is manifested by consuming more power, i.e., it reduces the economy of system operation, which is especially obvious when using the true load for optimization. Therefore, there exists a trade-off between the benefit and prediction horizon, which is 16 hours for MPC and 20 hours for RL in this work.

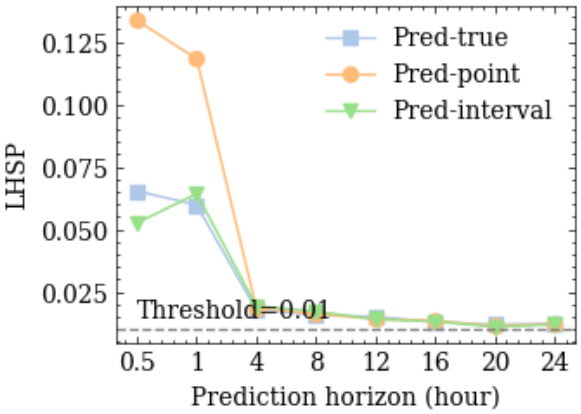
Morever, the results demonstrate that the uncertainty information of predictive load, provided by interval prediction, is much more important for MPC than RL. In both S_B and S_C, the combination of stochastic programming and interval prediction of load helps MPC to achieve lower LHSP than point prediction and even the true load, which makes it the only combination in MPC that successfully meets the LHSP constraint, despite a small loss in COE. However, the interval prediction seems to bring too much noise to RL, which makes it perform worse than point prediction. Since the RL is able to learn the relationship between input features and final targets, effective references close to true load may work better.



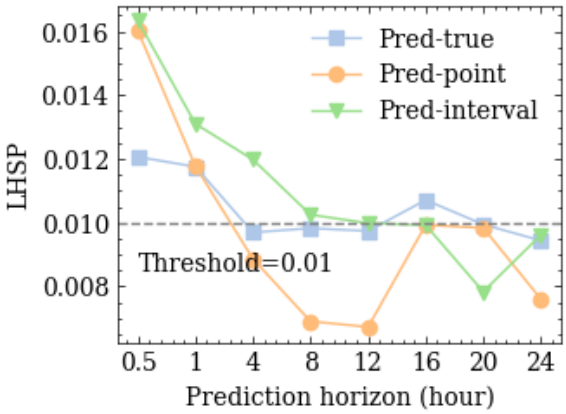
(a) LHSP of S_{B2} (MPC)



(b) LHSP of S_{B3} (RL)

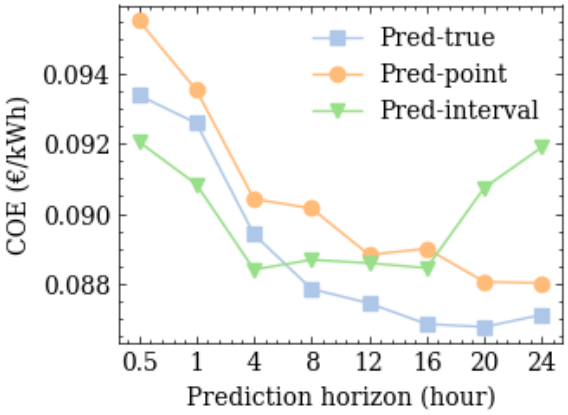


(c) LHSP of S_{C2} (MPC)

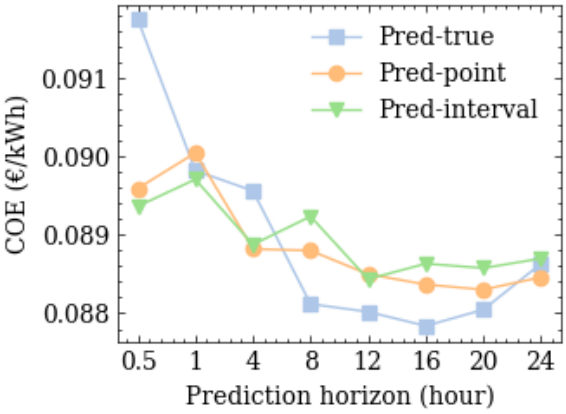


(d) LHSP of S_{C3} (RL)

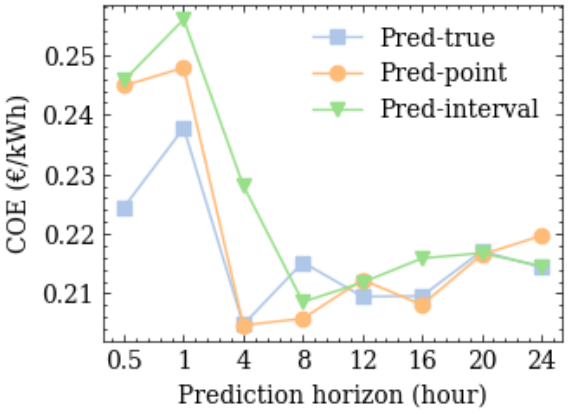
Figure 10 – Optimal LHSP under different prediction horizon in S_B and S_C.



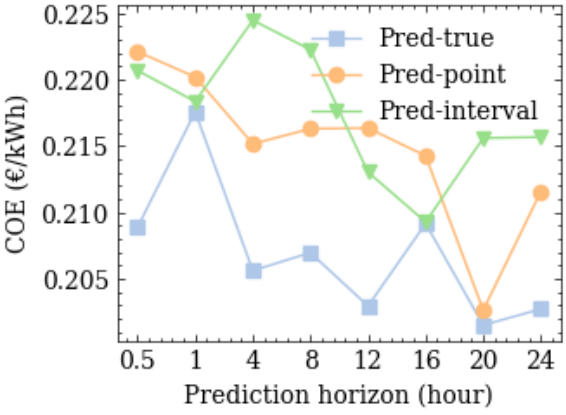
(a) COE of S_{B2} (MPC)



(b) COE of S_{B3} (RL)



(c) COE of S_{C2} (MPC)



(d) COE of S_{C3} (RL)

Figure 11 – Optimal COE under different prediction horizon in S_B and S_C.

6.5 Results of performance between optimization methods

In this work, swarm intelligence and data-driven optimization methods have been tested for capacity optimization. These two kinds of methods share many advantages, such as independence from the form of the optimization problem and the type of variables, and ease of parallel solving, which makes them more competitive than traditional methods for solving MINLP problems. However, they also have some shortcomings in terms of low computational efficiency, randomness and prematurity.

The differences in the ideas behind each method mean that their performance is also different. Table 16 compares their performance in the context of this work, in which PSO and BO share the same calculation environment settings. Generally, these two methods perform similarly in finding optimal solutions, with very small differences between the mean and the best values of their optimization results, even though BO provides the best solutions for both S_B and S_C. While BO costs about 6 times of PSO in computation time. The results reveal that PSO has higher computational efficiency but is more likely to fall into a local optimum, as compared to BO, which highlights the trade-off between efficiency and accuracy when choosing a suitable method. Furthermore, the combination of these two kinds of methods to improve the efficiency and accuracy of computationally costly optimization problems, such as co-optimization of capacity and operation, may be a better choice [47].

Table 16 – Comparison of performance between PSO and BO in capacity optimization of S_B and S_C.

* of 10 runs	S_{B0}	S_{B1}				S_{C1}			
	Best = Mean	Mean		Best		Mean		Best	
	PSO BO	PSO	BO	PSO	BO	PSO	BO	PSO	BO
COE(€/kWh)	0.094016	0.091868	0.091896	0.091863	0.091857	0.22965	0.22931	0.22137	0.22100
Better one	=	√	+2.8e-5	+6e-6	√	+3.4e-4	√	+3.7e-4	√
Standard deviation of COE	0	6.11e-6	40.47e-6			9.93e-3	8.88e-3		
Better one	=	√					√		
Computation time (10 runs, hour)	~0.4 converged	0.86	4.91	BO/PSO=5.7		4.09	26.03	BO/PSO=6.4	
Better one	~	√				√			

Based on the best solutions from capacity optimization, the MPC and RL have been introduced to optimize the operation. Table 17 summarises the best performance of MPC and RL in both S_B and S_C. From the aspects of optimization ability and computation time for online control, RL outperforms MPC. Since the built training environment of RL is closer to the real energy system than MPC, especially when there are

a large number of complex energy equipments, the mathematical planning method requires a lot of linear simplifications, which makes the optimization results deviate more from the real scenario requirements. In addition, when performing control, RL only requires a simple matrix operation to output the control strategy, while MP requires a more time-consuming solution to the full optimization problem. However, there exists uncertainty of the generalization ability of RL, in which the trained RL model fails to work on the test environment (raw time order load profiles) in S_B, but performs well when testing in S_C. The MPC requires uncertainty information of predictive load as inputs to meet the LHSP constraint. In contrast, RL is able to resolve the uncertainty of predictive load during the training process, which makes it perform better on point prediction of load. In general, the results indicate a better choice of MPC for the simple scenario and RL for the complex scenario. However, if the combination of MPC and RL can be achieved, it may offer more efficient and reliable control strategies in practice [12].

Table 17 – Comparison of best performance between MPC and RL in operation optimization of S_B and S_C.

Scenario	S_B		S_C	
Methods	MPC	RL	MPC	RL
LHSP constraint	√	√	×(0.0110)	√
COE (€/kWh)	0.088442	0.088020	(0.21675)	0.20264
Prediction horizon (hour)	16	20	20	20
Load prediction type	Interval	Point	Interval	Point
LHSP of raw load	√	√	-	√
COE of raw load	0.088447	0.098773	-	0.20263
Generalization	√	×	-	√
Computation time (hour) (single run given prediction horizon)	2~3	1~2 (Training)	4~30	3~12 (Training)

7. Conclusion

By comparing the optimal results of grid-based AEC-PEMFC-CHP system S_C with the HP system S_B, this paper reveals the former solution provides greater flexibility, albeit at the cost of higher electricity consumption and lower efficiency. The overall COE of S_C is 2.3 times that of S_B, requiring a 0.6 times increase in system efficiency and a reduction in equipment cost to 1/10 for S_C to be competitive with the best S_B solution.

The flexibility of both solutions relies heavily on the operation strategies. Based on the optimal capacity

solutions produced by PSO and BO, which have similar performance, MPC with interval prediction of load provides good optimization and generalization while meeting the reliability constraints. RL needs only point prediction of load to perform strong optimal ability, since RL is able to resolve the uncertainty of load during the training process. However, there exists a risk for the generalization of the trained RL model. Finally, in terms of prediction horizon of load, it's commonly found that a longer horizon, namely 16 and 20 hours, yields better performance.

With the help of advanced optimization methods, the tremendous flexibility potential offered by systems such as the EC-FC-CHP object of this study allows for a deep decoupling and matching of electricity-hydrogen-heat between supply and demand. This, in turn, allows for a more effective uptake of highly fluctuating but increasingly cost-effective RE power. Meeting the ambitious net-zero targets will require that increasingly large shares of RE power generation assets are deployed, at an increasingly fast pace. However, optimization solutions such as those explored in this paper have the potential to ease the transition to efficient low-carbon, demand-response oriented energy systems.

Acknowledgements

The authors gratefully acknowledge support from the Basic Science Center Program for Ordered Energy Conversion of the National Natural Science Foundation of China [Grant No. 51888103] and UK Research and Innovation through the Centre for Research into Energy Demand Solutions [Grant No. EP/R035288/1].

Appendix A. Tables and figures

Table A - 1 – Evaluation criteria for interval and point prediction.

Criteria	Equation	Value range
PB [37]	$\mathrm{PB}(\hat{y}_{t,q}, y_t, q) = \begin{cases} (1-q)(\hat{y}_{t,q} - y_t), & y_t < \hat{y}_{t,q} \\ 0, & y_t = \hat{y}_{t,q} \text{ (test the reliability/error)} \\ q(y_t - \hat{y}_{t,q}), & y_t > \hat{y}_{t,q} \end{cases}$ where: $\cdot q$ denotes the quantile value for prediction, $\cdot \hat{y}_{t,q}, y_t$ denote the predicted and true value at time step t.	$\mathrm{PB} \in [0, +\infty)$ Lower value means higher prediction reliability (i.e. lower prediction error).

WS [48]	$\mathrm{WS} = \begin{cases} \delta_t, & L_t \le y_t \le U_t \\ \delta_t + \frac{2(L_t - y_t)}{\alpha}, & y_t < L_t \\ \delta_t + \frac{2(y_t - U_t)}{\alpha}, & y_t > U_t \end{cases}$ (test the sharpness/interval size) where: $\cdot \delta_t = U_t - L_t$ denotes the length of prediction interval at time step t, $\cdot U_t$ and L_t denote the upper and lower boundaries of interval at time step t, $\cdot y_t$ denotes the true value at time step t $\cdot \alpha$ denotes the share of interval beyong prediction. It equals to $1 - 0.9 = 0.1$ here.	$\mathrm{WS} \in [0, +\infty)$ Lower value means higher prediction sharpness (i.e. lower interval size).
MAPE	$\mathrm{MAPE} = \frac{1}{T}\sum_{t=1}^{T} \frac{\lvert y_t - \hat{y}_t \rvert}{\lvert y_t \rvert} \times 100\%$ (test the accuracy) where: $\cdot T$ is the number of time steps, $\cdot \hat{y}_t, y_t$ denote the predicted and true value at time step t.	$\mathrm{MAPE} \in [0, +\infty)$ Lower value means higher prediction accuracy.
RMSE	$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}$ (test the error) where: $\cdot T$ is the number of time steps, $\cdot \hat{y}_t, y_t$ denote the predicted and true value at time step t.	$\mathrm{RMSE} \in [0, +\infty)$ Lower value means lower prediction error.

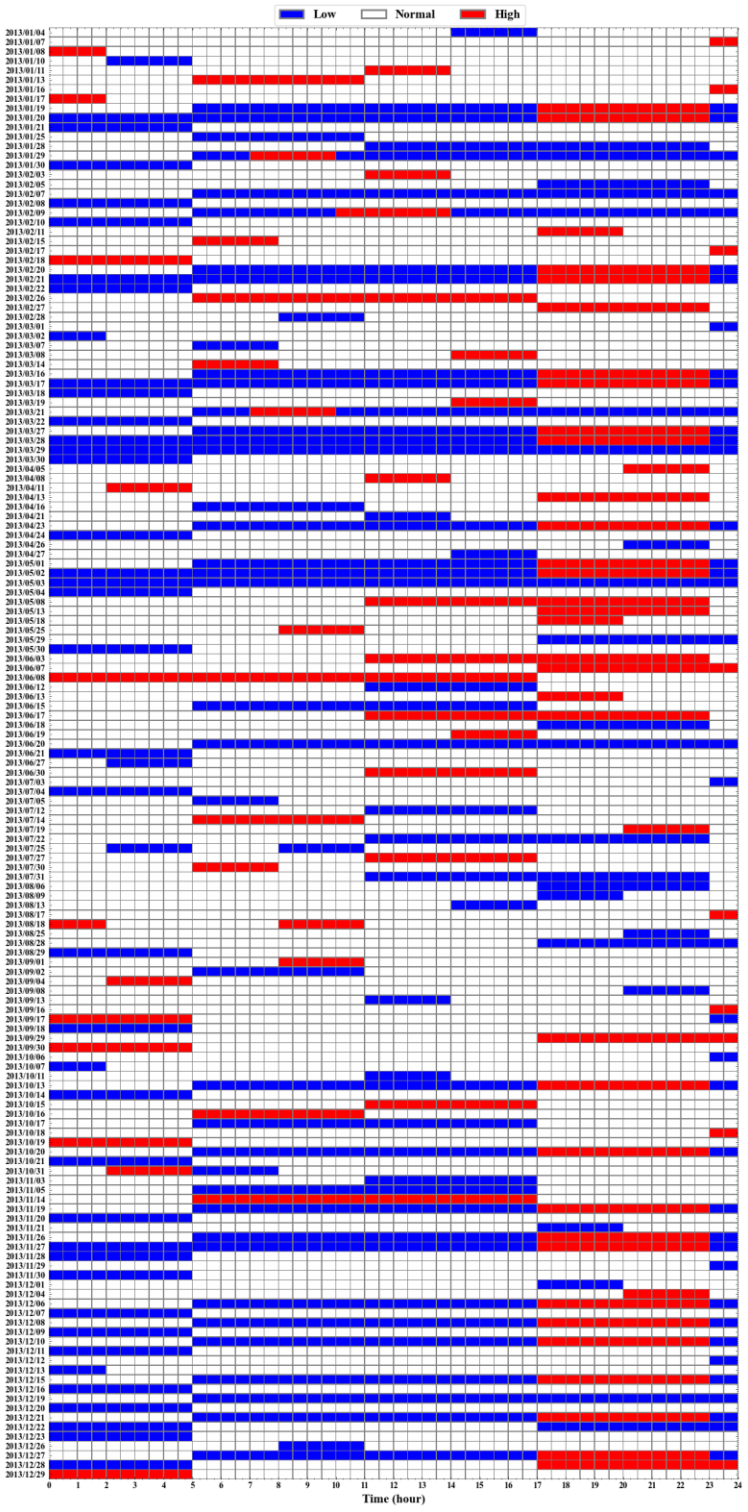
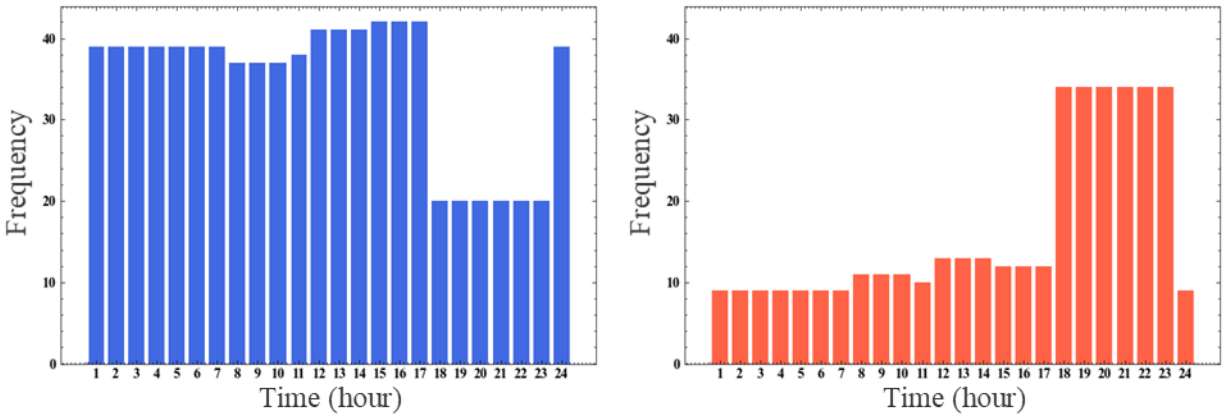


Figure A - 1 – Temporal distribution of dynamic Time-of-Use tariffs in 2013.



(a) Low tariff (b) High tariff

Figure A - 2 – Distribution of electricity dToU tariffs by the hour.

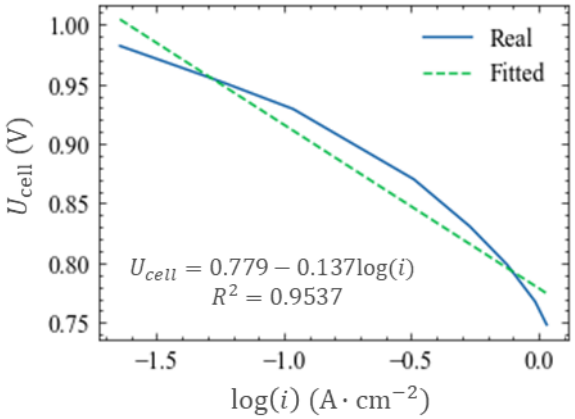


Figure A - 3 – Fitted Tafel equation from experimental data [25].

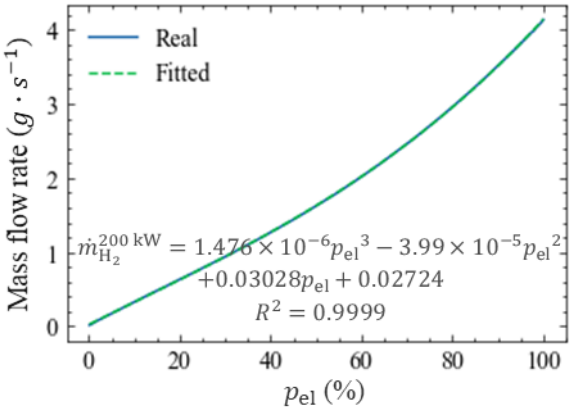


Figure A - 4 – Fitted relationship between hydrogen consumption and the percentage of output electrical power to rated electrical power from experimental data of 200 kW PEMFC [25].

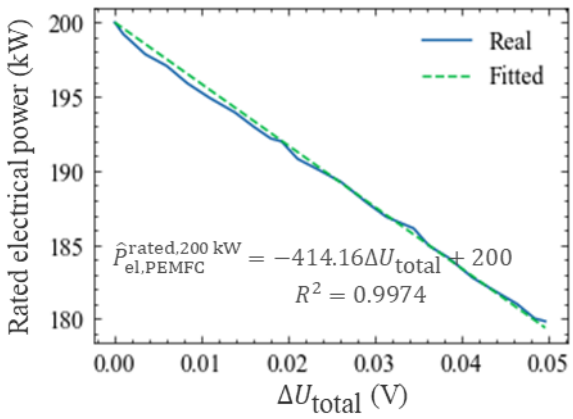
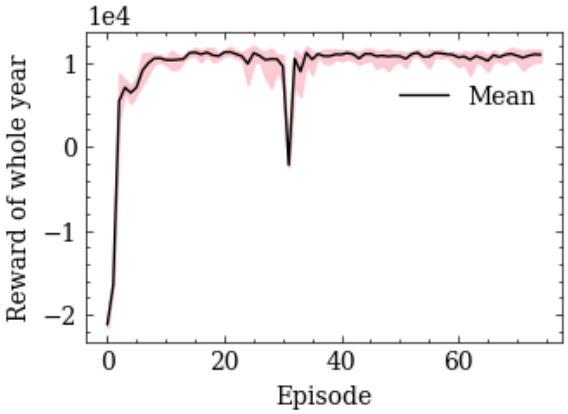
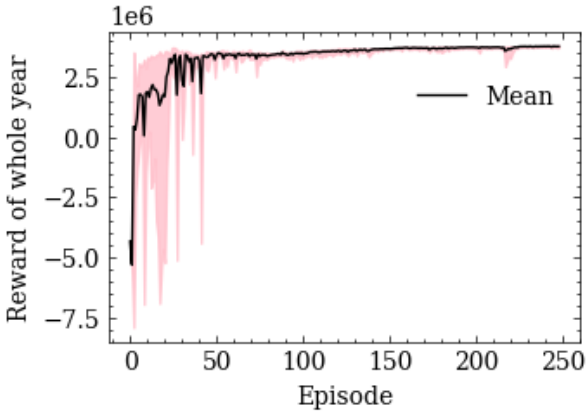


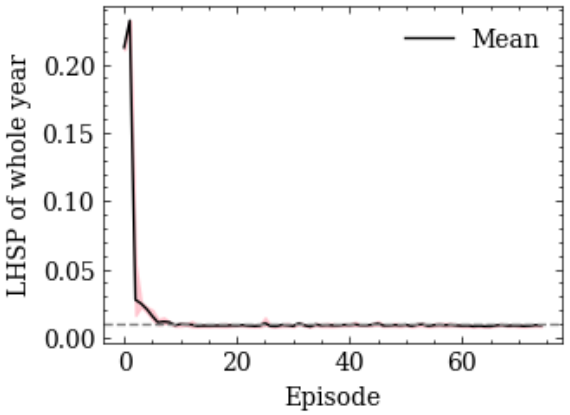
Figure A - 5 – Fitted relationship between rated electrical power and cell voltage loss from experimental data of 200 kW PEMFC [25].



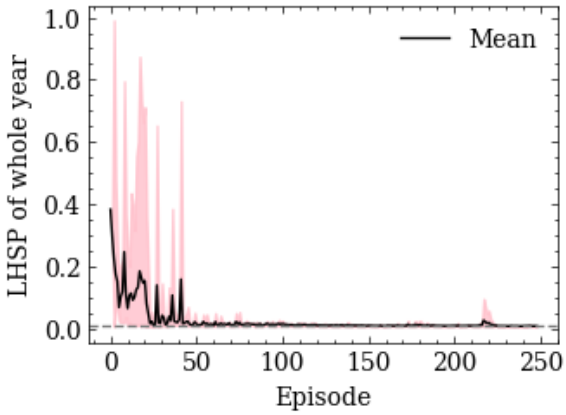
(a) Reward of S_{B3}



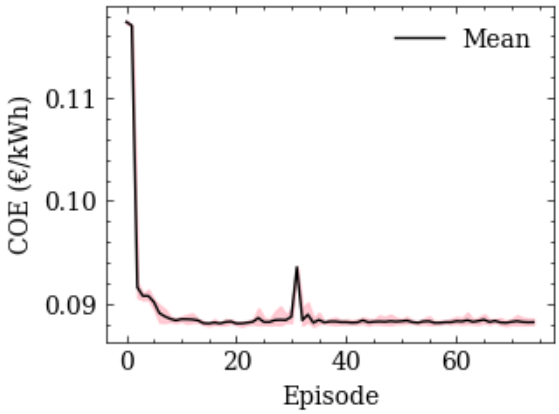
(b) Reward of S_{C3}



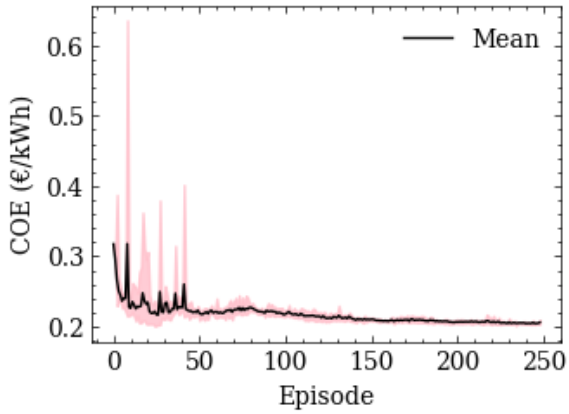
(c) LHSP of S_{B3}



(d) LHSP of S_{C3}

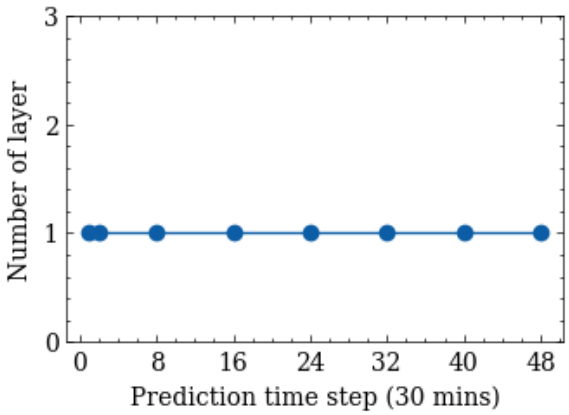


(e) COE of S_{B3}

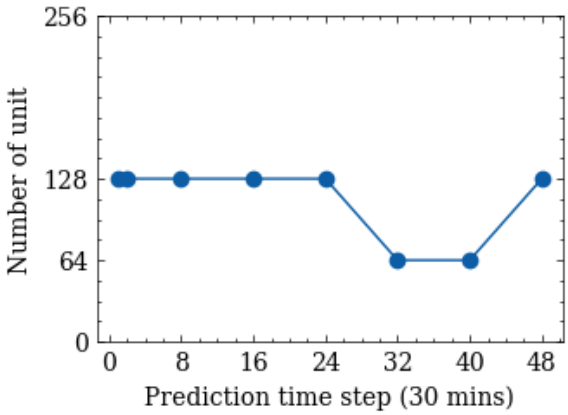


(f) COE of S_{C3}

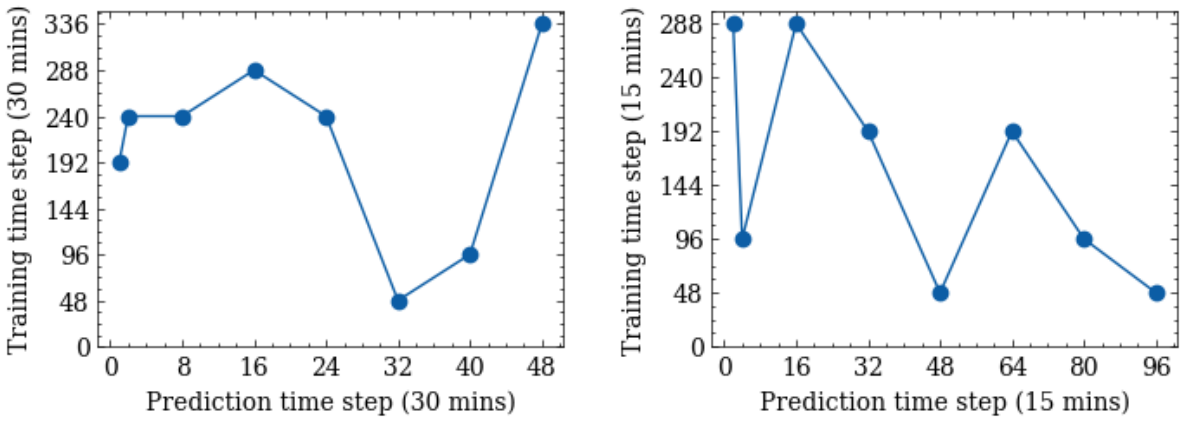
Figure A - 6 – Samples of RL training process using load from point prediction with prediction horizon of 20 hours (5 runs).



(a) Optimal layer number (EL)



(b) Optimal unit number (EL)



(c) Optimal training time steps (EL)

(d) Optimal training time steps (HL)

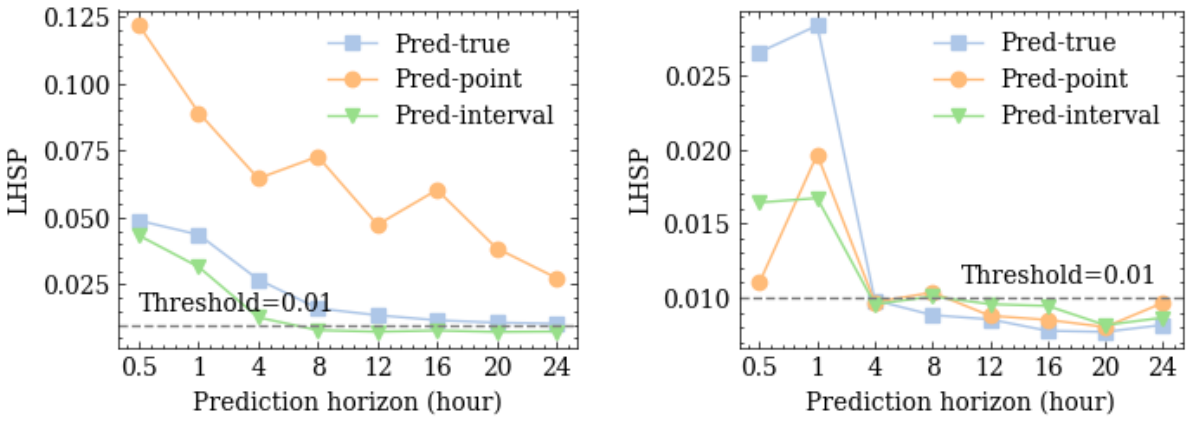
Figure A - 7 – Optimization results of hyperparameters for QLSTM model under different prediction time steps. (Electricity load (EL) prediction: (a), (b), (c); heat load (HL) prediction: (d))

Table A - 2 – Evaluation results of trained QLSTM models on test dataset for electricity load.

Prediction horizon (hour)	Mean PB (kWh)	Mean WS (kWh)	MAPE (%)	RMSE (kWh)
0.5	0.83	17.63	6.84	4.18
1	1.73	17.58	6.93	4.60
4	10.72	28.09	10.78	6.83
8	23.86	32.31	11.83	7.37
12	41.69	37.38	14.17	8.52
16	49.10	29.38	13.39	7.91
20	66.61	33.19	14.57	8.27
24	78.63	33.70	14.02	8.02

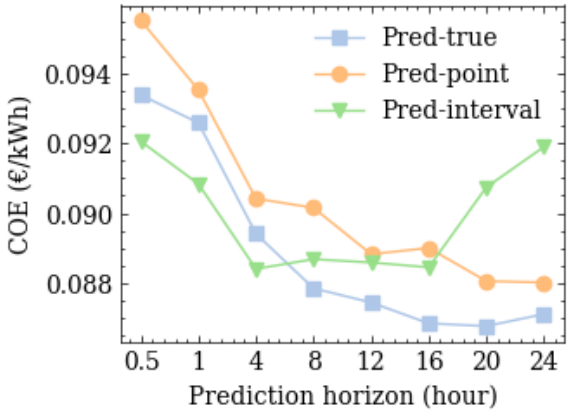
Table A - 3 – Evaluation results of trained QLSTM models on test dataset for heat load.

Prediction horizon (hour)	Mean PB (kWh)	Mean WS (kWh)	MAPE (%)	RMSE (kWh)
0.5	6.99	77.45	19.23	17.85
1	14.76	84.64	22.18	17.95
4	67.44	94.01	26.12	20.69
8	134.60	94.06	26.28	20.61
12	214.35	102.71	27.08	21.28
16	283.61	98.56	26.79	21.72
20	351.89	98.11	27.87	21.40
24	445.67	107.27	28.53	22.08

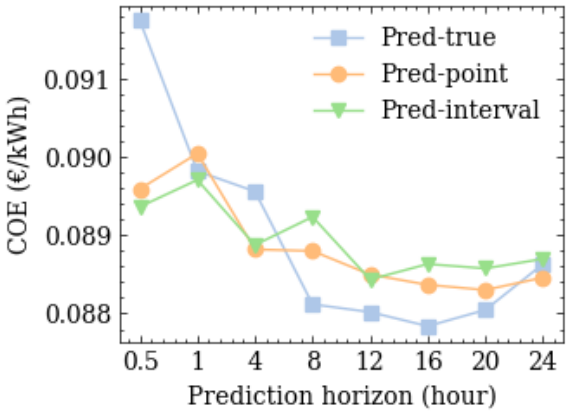


(a) LHSP of S_{B2} (MPC)

(b) LHSP of S_{B3} (RL)



(c) COE of S_{B2} (MPC)



(d) COE of S_{B3} (RL)

Figure A - 8 – Results of MPC and RL in S_B.

Table A - 4 – LHSP of S_{B2} using MPC

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.04883	0.12174	0.04326
1	0.04357	0.08916	0.03152
4	0.02681	0.06460	0.01268
8	0.01605	0.07278	**0.00789**
12	0.01350	0.04730	0.00734
16	0.01163	0.06032	0.00777
20	0.01071	0.03853	0.00725
24	0.01035	0.02746	0.00741

Table A - 5 – COE of S_{B2} using MPC

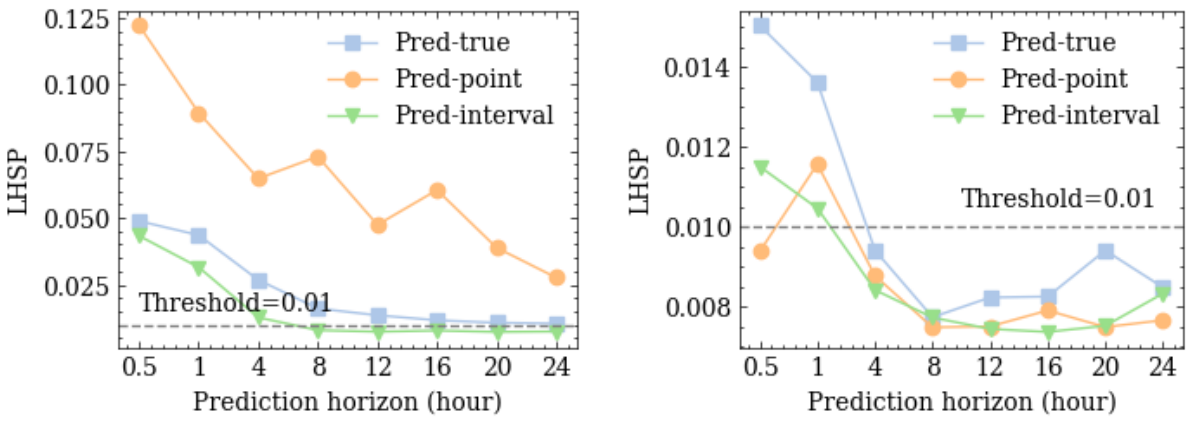
Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.093393	0.095540	0.092060
1	0.092600	0.093500	0.090800
4	0.089406	0.090417	0.088398
8	0.08784	0.090151	0.088676
12	0.087432	0.088820	0.088581
16	0.086833	0.088998	**0.088442**
20	0.086752	0.088046	0.090741
24	0.087099	0.088008	0.091902

Table A - 6 – LHSP of S_{B3} using RL (mean of 5 runs)

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.02659	0.01103	0.01643
1	0.02839	0.01967	0.01672
4	**0.00982**	**0.00969**	**0.00955**
8	0.00882	0.01035	0.01008
12	0.00853	0.00877	0.00956
16	0.00776	0.00850	0.00944
20	0.00768	0.00801	0.00814
24	0.00815	0.00964	0.00864

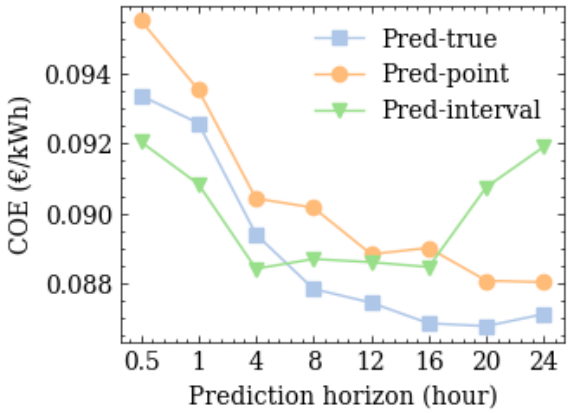
Table A - 7 – COE of S_{B3} using RL (mean of 5 runs)

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.091744	0.089591	0.089363
1	0.089818	0.090044	0.089707
4	0.089560	0.088819	0.088875
8	0.088118	0.088799	0.089235
12	0.088019	0.088499	**0.088435**
16	**0.087839**	0.088367	0.088632
20	0.088046	**0.088301**	0.088576
24	0.088631	0.088459	0.088697

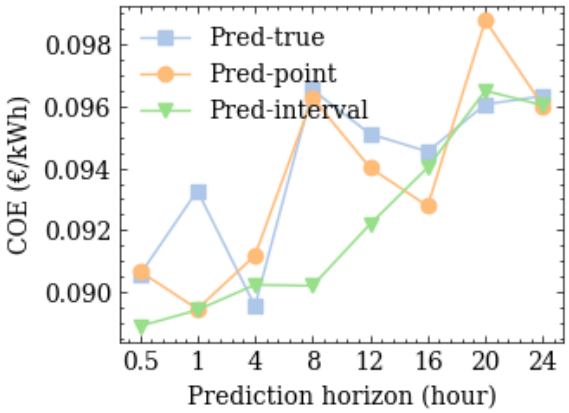


(a) LHSP of raw S_{B2} (MPC)

(b) LHSP of raw S_{B3} (RL)



(c) COE of raw S_{B2} (MPC)



(d) COE of raw S_{B3} (RL)

Figure A - 9 – Evaluation of generalization for MPC and RL in S_B on raw time order load profiles.

Table A - 8 – LHSP of S_{B2} using MPC on raw time order load profiles

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.04882	0.12217	0.04327
1	0.04354	0.08943	0.03150
4	0.02676	0.06479	0.01268
8	0.01604	0.07302	**0.00789**
12	0.01350	0.04743	0.00734
16	0.01163	0.06046	0.00777
20	0.01069	0.03884	0.00725
24	0.01038	0.02763	0.00741

Table A - 9 – COE of S_{B2} using MPC on raw time order load profiles

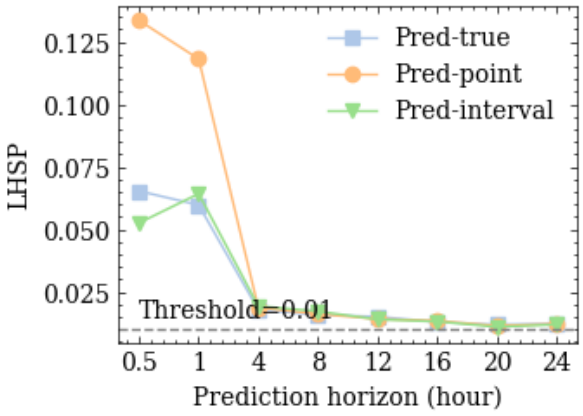
Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.093385	0.095549	0.092055
1	0.092589	0.093551	0.090836
4	0.089390	0.090422	0.088399
8	0.087816	0.090156	0.088678
12	0.087421	0.088818	0.088587
16	0.086827	0.088999	**0.088447**
20	0.086745	0.088054	0.090748
24	0.087094	0.088012	0.091911

Table A - 10 – LHSP of S_{B3} using RL on raw time order load profiles (based on best RL models)

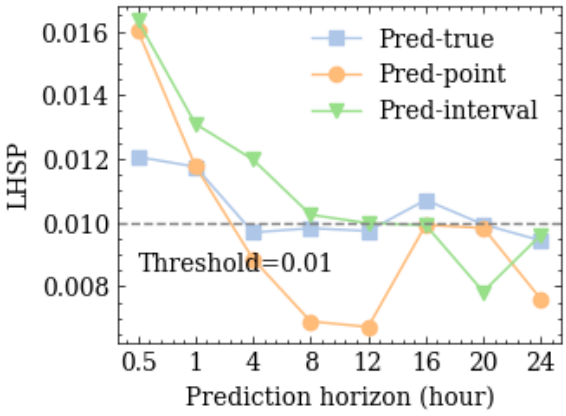
Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.01504	0.00939	0.01151
1	0.01362	0.01157	0.01047
4	**0.00941**	**0.00878**	**0.00842**
8	0.00774	0.00749	0.00773
12	0.00824	0.00751	0.00745
16	0.00826	0.00791	0.00737
20	0.00941	0.00750	0.00753
24	0.00849	0.00767	0.00833

Table A - 11 – COE of S_{B3} using RL on raw time order load profiles (based on best RL models)

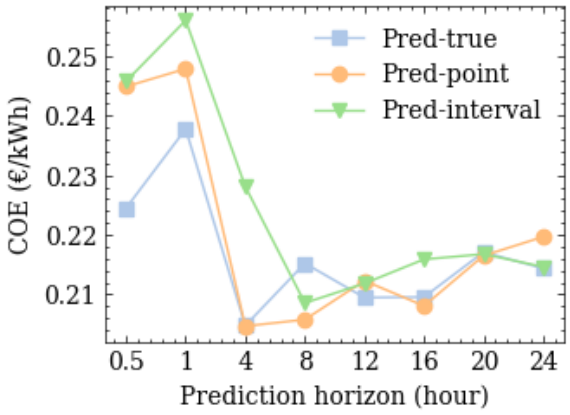
Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.090553	0.090672	0.088899
1	0.093278	0.089452	0.089420
4	**0.089549**	**0.091180**	0.090236
8	0.096528	0.096275	**0.090202**
12	0.095106	0.094012	0.092194
16	0.094536	0.092775	0.094023
20	0.096064	0.098773	0.096502
24	0.096330	0.096021	0.096032



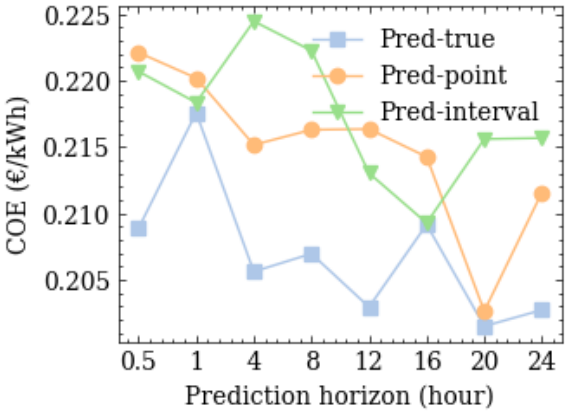
(a) LHSP of S_{C2} (MPC)



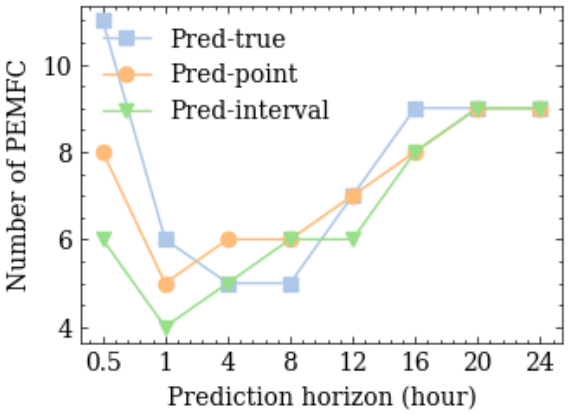
(b) LHSP of S_{C3} (RL)



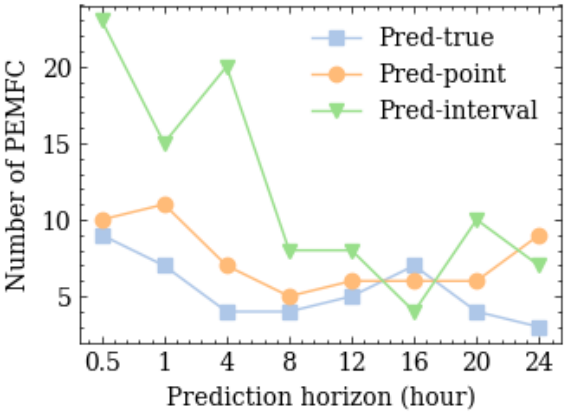
(c) COE of S_{C2} (MPC)



(d) COE of S_{C3} (RL)



(e) PEMFC number of S_{C2} (MPC)



(f) PEMFC number of S_{C3} (RL)

Figure A - 10 – Results of MPC and RL in S_C.

Table A - 12 – LHSP of S_{C2} using MPC

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.06534	0.1335	0.05262
1	0.05980	0.11820	0.06430
4	0.01788	0.01873	0.01935
8	0.01588	0.01619	0.01726
12	0.01514	0.01423	0.01403
16	0.01301	0.01346	0.01319
20	0.01195	0.01151	0.01103
24	0.01238	0.01222	0.01197

Table A - 13 – COE of S_{C2} using MPC

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.22449	0.24488	0.24578
1	0.23780	0.24790	0.25600
4	0.20474	0.20466	0.22810
8	0.21520	0.20576	0.20855
12	0.20946	0.21228	0.21182
16	0.20954	0.20806	0.21588
20	0.21708	0.21649	0.21675
24	0.21435	0.21965	0.21461

Table A - 14 – PEMFC number of S_{C2} using MPC

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	11	8	6
1	6	5	4
4	5	6	5
8	5	6	6
12	7	7	6
16	9	8	8
20	9	9	9
24	9	9	9

Table A - 15 – LHSP of S_{C3} using RL (mean of 5 runs)

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.01206	0.01603	0.01634
1	0.01175	0.01178	0.01310
4	**0.00969**	**0.00886**	0.01198
8	0.00982	0.00690	0.01025
12	0.00974	0.00673	**0.00998**
16	0.01071	0.00992	0.00992

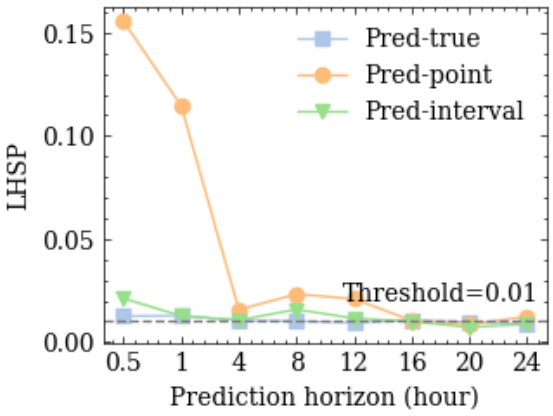
20	0.00994	0.00992	0.00783
24	0.00944	0.00760	0.00961

Table A - 16 – COE of S_{C3} using RL (mean of 5 runs)

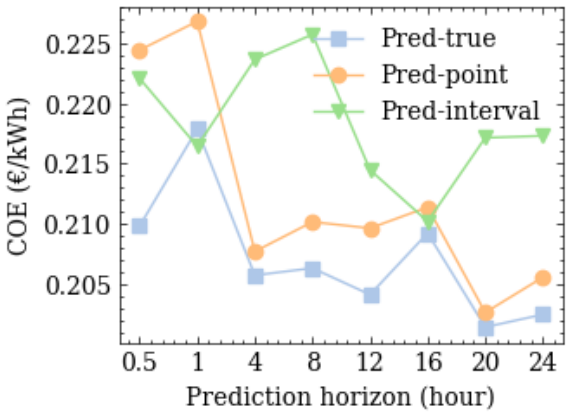
Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.20889	0.22212	0.22069
1	0.21759	0.22020	0.21835
4	0.20562	0.21517	0.22448
8	0.20697	0.21630	0.22226
12	0.20296	0.21636	0.21304
16	0.20917	0.21426	**0.20928**
20	**0.20151**	**0.20620**	0.21560
24	0.20273	0.21157	0.21568

Table A - 17 – PEMFC number of S_{C3} using RL (mean of 5 runs)

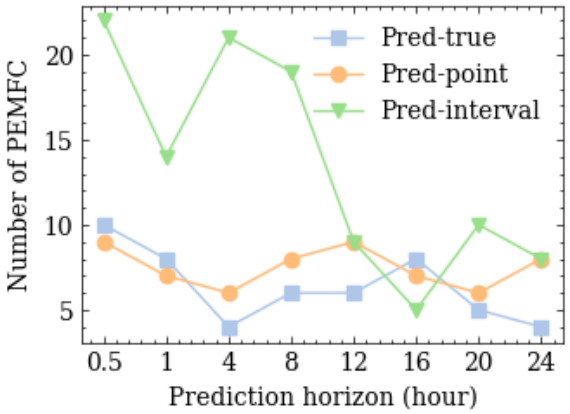
Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	9	10	23
1	7	11	15
4	4	7	20
8	4	5	8
12	5	6	8
16	7	6	**4**
20	**4**	**6**	10
24	3	9	7



(a) LHSP of raw S_{C3} (RL)



(b) COE of raw S_{C3} (RL)



(c) PEMFC number of raw S_{C3} (RL)

Figure A - 11 – Evaluation of generalization for RL in S_C on raw time order load profiles.

Table A - 18 – LHSP of S_{C3} using RL on raw time order load profiles (based on best RL models)

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.01254	0.15549	0.02112
1	0.01277	0.11448	0.01263
4	0.01045	0.01570	0.01060
8	0.01011	0.02319	0.01576
12	**0.00932**	0.02090	0.01149
16	0.01058	0.01041	**0.00987**
20	0.00958	**0.00841**	0.00719
24	0.00905	0.01235	0.00861

Table A - 19 – COE of S_{C3} using RL on raw time order load profiles (based on best RL models)

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval
0.5	0.20981	0.22442	0.22207
1	0.21786	0.22680	0.21644
4	0.20572	0.20770	0.22364
8	0.20632	0.21015	0.22572
12	0.20412	0.20966	0.21446
16	0.20922	0.21138	**0.21020**
20	**0.20141**	**0.20263**	0.21714
24	0.20246	0.20555	0.21730

Table A - 20 – PEMFC number of S_{C3} using RL on raw time order load profiles (based on best RL models)

Prediction horizon (hour)	Pred-true	Pred-point	Pred-interval

0.5	10	9	22
1	8	7	14
4	4	6	21
8	6	8	19
12	6	9	9
16	8	7	**5**
20	**5**	**6**	10
24	4	8	8

Appendix B. Description of optimization problems

B.1 Optimization problem of MPC

Following minimization form, the objective function of S_{B2} is expressed as follows:

$$\mathrm{J}_{\mathrm{S_{B2}}} = \mathrm{J}_{\mathrm{source}}(P_{\mathrm{HP}}) + \mathrm{J}_{\mathrm{storage}}(Q_{\mathrm{HWT,discharge}}, Q_{\mathrm{HWT,charge}}) - \mathrm{J}_{\mathrm{load}}(P_{\mathrm{HP}}) \quad (\text{B.1-1})$$

where

$$\mathrm{J}_{\mathrm{source}}(P_{\mathrm{HP}}) = \sum_{t}^{\mathrm{T_{fut}}} \mathrm{c}_{\mathrm{e}}^{t} P_{\mathrm{HP}}^{t} \Delta \mathrm{t} \quad (\text{B.1-2})$$

$$\mathrm{J}_{\mathrm{storage}}(Q_{\mathrm{HWT,discharge}}, Q_{\mathrm{HWT,charge}}) = \mathrm{J}_{\mathrm{HWT}}(Q_{\mathrm{HWT,discharge}}) + \mathrm{J}_{\mathrm{HWT}}(Q_{\mathrm{HWT,discharge}}, Q_{\mathrm{HWT,charge}}) \quad (\text{B.1-3})$$

$$\mathrm{J}_{\mathrm{load}}(P_{\mathrm{HP}}) = \begin{cases} \sum_{t}^{\mathrm{T_{fut}}} -\mathrm{c}_{\mathrm{e}}^{t} W_{\mathrm{HP}}^{t}, & \text{if } \mathrm{q}_{\widehat{\mathrm{W}}_{\mathrm{EL}}}^{t} \geq 95\% \\ \sum_{t}^{\mathrm{T_{fut}}} \mathrm{c}_{\mathrm{e}}^{t} \mathrm{W}_{\mathrm{HP}}^{t}, & \text{if } \mathrm{q}_{\widehat{\mathrm{W}}_{\mathrm{EL}}}^{t} \leq 5\% \end{cases} \quad (\text{B.1-4})$$

For cost from storage $\mathrm{J}_{\mathrm{storage}}(Q_{\mathrm{HWT,discharge}}, Q_{\mathrm{HWT,charge}})$, it contains heat discharging from hot water tank and a penalty for using high price heat, which are expressed as follows:

$$\mathrm{J}_{\mathrm{storage}}(Q_{\mathrm{HWT,discharge}}) = \sum_{t}^{\mathrm{T_{fut}}} \mathrm{c}_{\mathrm{h,HWT,discharge}}^{t} Q_{\mathrm{HWT,discharge}}^{t} \quad (\text{B.1-5})$$

$$\mathrm{J}_{\mathrm{storage}}(Q_{\mathrm{HWT,discharge}}, Q_{\mathrm{HWT,charge}}) = \sum_{t}^{\mathrm{T_{fut}}} \left(\frac{\mathrm{c}_{\mathrm{e}}^{t}}{\mathrm{COP}^{t}} - \mathrm{c}_{\mathrm{h,HWT,discharge}}^{t}\right)\left(Q_{\mathrm{HWT,charge}}^{t} - Q_{\mathrm{HWT,discharge}}^{t}\right) \quad (\text{B.1-6})$$

The constraints of S_{B2} include:

$$\mathrm{LHSP} = \frac{\sum_{t}^{\mathrm{T_{fut}}} Q_{\mathrm{HL}}^{t} - P_{\mathrm{HP}}^{t}\Delta \mathrm{tCOP}^{t} + Q_{\mathrm{HWT,charge}}^{t} - Q_{\mathrm{HWT,discharge}}^{t}}{\sum_{t}^{\mathrm{T_{fut}}} Q_{\mathrm{HL}}^{t}} \leq 0.01 \quad (\text{B.1-7})$$

$$\mathrm{SoC}_{\mathrm{HWT}}^{t} = \mathrm{SoC}_{\mathrm{HWT}}^{\mathrm{ini}}, t = 1 \quad (\text{B.1-8})$$

$$\mathrm{SoC}_{\mathrm{HWT}}^{t} = (\mathrm{SoC}_{\mathrm{HWT}}^{t-1} + \frac{Q_{\mathrm{HWT,charge}}^{t} - Q_{\mathrm{HWT,discharge}}^{t}}{\mathrm{Q}_{\mathrm{HWT}}^{\mathrm{rated}}}) \times (1 - \eta_{\mathrm{loss,HWT}}), t = 2, \cdots, \mathrm{T_{fut}} \quad (\text{B.1-9})$$

$$Q_{\mathrm{HWT,discharge}}^{t} \times Q_{\mathrm{HWT,charge}}^{t} = 0, t = 1, \cdots, \mathrm{T_{fut}} \quad (\text{B.1-10})$$

In S_{C2}, the objective is expressed as:

$$\mathrm{J}_{S_{C2}} = \mathrm{J}_{\mathrm{source}}(P_{\mathrm{AEC}}, m_{\mathrm{HT,charge}}, P_{\mathrm{el,PEMFC}}) + \mathrm{J}_{\mathrm{storage}}(m_{\mathrm{HT,discharge}}, Q_{\mathrm{HWT,discharge}}) - \mathrm{J}_{\mathrm{load}}(P_{\mathrm{el,PEMEC}}, P_{\mathrm{AEC}}) + \mathrm{J}_{\mathrm{life}}(P_{\mathrm{el,PEMFC}}) \quad (\text{B.1-11})$$

Where

$$\mathrm{J}_{\mathrm{source}}(P_{\mathrm{AEC}}, m_{\mathrm{HT,charge}}, P_{\mathrm{el,PEMFC}}) = \sum_{t}^{\mathrm{T_{fut}}} \mathrm{c}_{\mathrm{e}}^{t}(P_{\mathrm{AEC}}^{t}\Delta \mathrm{t}/\eta_{\mathrm{AC/DC}} + \mathrm{w}_{\mathrm{compressor}} m_{\mathrm{HT,charge}}^{t} - \eta_{\mathrm{DC/AC}} P_{\mathrm{el,PEMFC}}^{t}\Delta \mathrm{t}) \quad (\text{B.1-12})$$

$$\mathrm{J}_{\mathrm{storage}}(m_{\mathrm{HT,discharge}}, Q_{\mathrm{HWT,discharge}}) = \mathrm{J}_{\mathrm{HT}}(m_{\mathrm{HT,discharge}}) + \mathrm{J}_{\mathrm{HWT}}(Q_{\mathrm{HWT,discharge}}) \quad (\text{B.1-13})$$

$$\mathrm{J}_{\mathrm{load}}(P_{\mathrm{el,PEMEC}}, P_{\mathrm{AEC}}) = \begin{cases} \sum_{t}^{\mathrm{T_{fut}}} (P_{\mathrm{el,PEMFC}}^{t} - P_{\mathrm{AEC}}^{t})\Delta \mathrm{t}, & \text{if } \mathrm{q}_{\widehat{\mathrm{W}}_{\mathrm{EL}}}^{t} \geq 95\% \\ \sum_{t}^{\mathrm{T_{fut}}} -(P_{\mathrm{el,PEMFC}}^{t} - P_{\mathrm{AEC}}^{t})\Delta \mathrm{t}, & \text{if } \mathrm{q}_{\widehat{\mathrm{W}}_{\mathrm{EL}}}^{t} \leq 5\% \end{cases} \quad (\text{B.1-14})$$

$$\mathrm{J}_{\mathrm{life}}(P_{\mathrm{el,PEMFC}}) = \frac{\Delta U_{\mathrm{total}}^{\mathrm{t=1}}}{\mathrm{U}_{\mathrm{cell}}^{\mathrm{rated}}(1 - \eta_{\mathrm{U}}^{\mathrm{thre}})} \mathrm{cost}_{\mathrm{PEMFC}}^{\mathrm{re}} \mathrm{P}_{\mathrm{el,PEMFC}}^{\mathrm{rated}} \quad (\text{B.1-15})$$

For cost from storage $\mathrm{J}_{\mathrm{storage}}(m_{\mathrm{HT,discharge}}, Q_{\mathrm{HWT,discharge}})$, it contains heat discharging from hot water which is the same as S_{B2} (B.1-5), and the hydrogen discharging from hydrogen tank which can be calculated as:

$$\mathrm{J}_{\mathrm{HT}}(m_{\mathrm{HT,discharge}}) = \sum_{t}^{\mathrm{T_{fut}}} \mathrm{c}_{\mathrm{H_2,HT,discharging}}^{t} m_{\mathrm{HT,discharge}}^{t} \quad (\text{B.1-16})$$

Besides, it's noted that only the voltage loss of next time step for control action of PEMFC will be considered due to the tradeoff between solving speed and effects.

The constraint functions of S_{C2} include:

$$\mathrm{LHSP}^{\mathrm{t}} = \frac{Q_{\mathrm{HL}}^{t} - Q_{\mathrm{PEMFC}}^{t} + Q_{\mathrm{HWT,charge}}^{t} - Q_{\mathrm{HWT,discharge}}^{t}}{Q_{\mathrm{HL}}^{t}} \leq 0.01, t = 1, \cdots, \mathrm{T_{fut}} \quad (\text{B.1-17})$$

$$\mathrm{SoC}_{\mathrm{HWT}}^{t} = \mathrm{SoC}_{\mathrm{HWT}}^{\mathrm{ini}}, t = 1 \quad (\text{B.1-18})$$

$$\mathrm{SoC}_{\mathrm{HWT}}^{t} = (\mathrm{SoC}_{\mathrm{HWT}}^{t-1} + \frac{Q_{\mathrm{HWT,charge}}^{t} - Q_{\mathrm{HWT,discharge}}^{t}}{\mathrm{Q}_{\mathrm{HWT}}^{\mathrm{rated}}}) \times (1 - \eta_{\mathrm{loss,HWT}}), t = 2, \cdots, \mathrm{T_{fut}} \quad (\text{B.1-19})$$

$$Q_{\mathrm{HWT,discharge}}^{t} \times Q_{\mathrm{HWT,charge}}^{t} = 0, t = 1, \cdots, \mathrm{T_{fut}} \quad (\text{B.1-20})$$

$$\mathrm{SoC}_{\mathrm{HT}}^{t} = \mathrm{SoC}_{\mathrm{HT}}^{\mathrm{ini}}, t = 1 \quad (\text{B.1-21})$$

$$\mathrm{SoC}_{\mathrm{HT}}^{t} = (\mathrm{SoC}_{\mathrm{HT}}^{t-1} + \frac{m_{\mathrm{HT,charge}}^{t} - m_{\mathrm{HT,discharge}}^{t}}{\mathrm{m}_{\mathrm{HT}}^{\mathrm{rated}}}), t = 2, \cdots, \mathrm{T_{fut}} \quad (\text{B.1-22})$$

$$m_{\mathrm{HT,discharge}}^{t} \times m_{\mathrm{HT,charge}}^{t} = 0, t = 1, \cdots, \mathrm{T_{fut}} \quad (\text{B.1-23})$$

While using predictive loads from interval prediction, the stochastic programming considering three sceneries with respect to quantiles $\mathrm{q} = \{0.05, 0.5, 0.95\}$ will be used by adding the constraints accordingly and producing a weighted objective function corresponding to weights $\mathrm{w}_q = \{0.05, 0.9, 0.05\}$, as follows:

$$\mathrm{J}_{\mathrm{S_{B2}}}^{\mathrm{q}} = \mathrm{J_{source}}(P_{\mathrm{HP}}) + \sum_q \mathrm{w}_q \, \mathrm{J}_{\mathrm{storage}}^{q}(Q_{\mathrm{HWT,discharge}}, Q_{\mathrm{HWT,charge}}) - \sum_q \mathrm{w}_q \, \mathrm{J}_{\mathrm{load}}^{q}(P_{\mathrm{HP}}) \quad (\text{B.1-24})$$

$$\mathrm{J}_{\mathrm{S_{C2}}}^{\mathrm{q}} = \sum_q \mathrm{w}_q \, \mathrm{J}_{\mathrm{source}}^{q}(P_{\mathrm{AEC}}, m_{\mathrm{HT,charge}}, P_{\mathrm{el,PEMFC}}) + \sum_q \mathrm{w}_q \, \mathrm{J}_{\mathrm{storage}}^{q}(m_{\mathrm{HT,discharge}}, Q_{\mathrm{HWT,discharge}}) - \sum_q \mathrm{w}_q \, \mathrm{J}_{\mathrm{load}}^{q}(P_{\mathrm{el,PEMEC}}, P_{\mathrm{AEC}}) + \mathrm{J_{life}}(P_{\mathrm{el,PEMFC}}) \quad (\text{B.1-25})$$

Appendix C. Methods

C.1 Quantile long-short term memory neural network (QSLTM)

Long-short term memory (LSTM) [38] improved from typical recurrent neural network (RNN) can effectively address vanishing gradients to capture long term dependencies by replacing the ordinary recurrent cell with the memory cell. Figure C - 1 shows the workflow of LSTM at time step t, in which the memory cell consists of an internal state $\mathrm{C}_{\mathrm{t-1}}$ and three multiplicative gates.

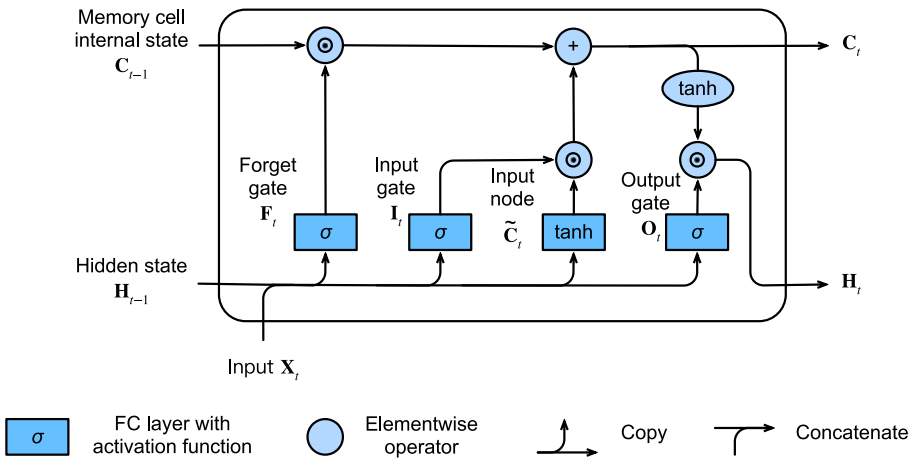


Figure C - 1 – Computing the hidden state in an LSTM model [49].

The forget gate $\mathrm{F_t}$ and input gate $\mathrm{I_t}$ are designed to control the update of $\mathrm{C_{t-1}}$ by deciding whether internal state should be flushed to 0 and then add the information from current input (candidate memory cell $\tilde{\mathbf{C}}_t$), respectively. The output gate $\mathrm{O_t}$ decides if the update of hidden state $\mathrm{H_{t-1}}$ considers the information from the updated internal cell $\mathrm{C_t}$. The three gates are calculated by:

$$\begin{aligned}\mathbf{I}_t &= \sigma(\mathbf{X}_t\mathbf{W}_{xi} + \mathbf{H}_{t-1}\mathbf{W}_{hi} + \mathbf{b}_i)\\ \mathbf{F}_t &= \sigma(\mathbf{X}_t\mathbf{W}_{xf} + \mathbf{H}_{t-1}\mathbf{W}_{hf} + \mathbf{b}_f)\\ \mathbf{O}_t &= \sigma(\mathbf{X}_t\mathbf{W}_{xo} + \mathbf{H}_{t-1}\mathbf{W}_{ho} + \mathbf{b}_o)\end{aligned} \quad (\text{C.1-1})$$

The candidate memory cell $\tilde{\mathbf{C}}_t$ represents current input information and is calculated by:

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{X}_t\mathbf{W}_{xc} + \mathbf{H}_{t-1}\mathbf{W}_{hc} + \mathbf{b}_c) \quad (\text{C.1-2})$$

where the $\mathbf{W}_{xi}, \mathbf{W}_{xf}, \mathbf{W}_{xo}, \mathbf{W}_{xc}$ and $\mathbf{W}_{hi}, \mathbf{W}_{hf}, \mathbf{W}_{ho}, \mathbf{W}_{hc}$ are weight matrix; the $\mathbf{b}_i, \mathbf{b}_f, \mathbf{b}_o, \mathbf{b}_c$ are bias parameters; the σ is sigmoid function with value limited to (0,1); the value of tanh function is limited to (-1,1).

The updated internal state $\mathrm{C_t}$ and hidden state $\mathrm{H_t}$ are calculated by:

$$\mathbf{C}_t = \mathbf{F}_t \odot \mathbf{C}_{t-1} + \mathbf{I}_t \odot \tilde{\mathbf{C}}_t \quad (\text{C.1-3})$$

$$\mathbf{H}_t = \mathbf{O}_t \odot \tanh(\mathbf{C}_t) \quad (\text{C.1-4})$$

where $\odot$ is the Hadamard (elementwise) product operator.

Finally, the output $\hat{\mathbf{Y}}_t$ of neural network is calculated by using a full connected layer with input of $\mathrm{H_t}$:

$$\hat{\mathbf{Y}}_t = \phi(\mathbf{H}\mathbf{W}_{hy} + \mathbf{b}_y) \quad (\text{C.1-5})$$

where $\mathbf{W}_{hy}$ is a weight matrix, and $\mathbf{b}_y$ is a bias parameter.

The parameters can be updated in training process using back-propagation with stochastic gradient descent. Here, the Pinball (PB) loss function in quantile regression [37] is introduced to achieve internal prediction, in which the PB is calculated by:

$$\mathrm{PB}(\hat{y}_{t,q}, y_t, q) = \begin{cases} (1-q)\left(\hat{y}_{t,q} - y_t\right), & y_t < \hat{y}_{t,q} \\ 0, & y_t = \hat{y}_{t,q} \\ q\left(y_t - \hat{y}_{t,q}\right), & y_t > \hat{y}_{t,q} \end{cases} \quad \text{(C.1-6)}$$

where, $q \in (0,1)$ is the quantile. When $q = 0.5$, PB is equal to the calculation of half mean absolute error (MAE). The quantile regression allows for estimation of given quantiles of y conditional on X without request to prespecify distribution for y.

In this work, 0.05, 0.5 and 0.95 are taken as the value of q, in which the lower and upper quantiles provide a 90% interval prediction and the median is chosen as the results representing point prediction.

C.2 Particle swarm optimization algorithm (PSO)

Particle swarm optimization (PSO) was first systematically proposed by Kennedy and Eberhart in 1995 [50], which does not depend on the form of the optimization problem and has the advantages of fast convergence and high optimization-seeking ability [51]. PSO simulates the foraging process of a flock of birds and finds the optimal solution by iteratively updating the velocity v and position x (i.e. solution) of each particle. In standard PSO, the v_{id}^{k+1} and x_{id}^{k+1} of dimension d in particle i updated from iteration k are calculated by:

$$\begin{cases} v_{id}^{k+1} = \omega \cdot v_{id}^{k+1} + c_1 r_1 \cdot \left(x_{id,\mathrm{pb}}^{k} - x_{id}^{k}\right) + c_2 r_2 \cdot \left(x_{id,\mathrm{gb}}^{k} - x_{id}^{k}\right) \\ x_{id}^{k+1} = x_{id}^{k} + v_{id}^{k+1} \end{cases} \quad \text{(C.2-1)}$$

where ω is inertia weight. Acceleration coefficients c_1, c_2 represent self-confidence and swarm confidence, respectively. r_1 and r_2 are random number sampling from U(0,1) uniform distribution. The subscripts pb and gb denote the personal optimum of particle i and global optimum of all particles, respectively.

The standard PSO is likely to achieve premature convergence in optimized functions with multiple local extremes due to rapid disappearance of the diversity of particles. To address this problem, this paper proposes an improved PSO by combining simplified PSO [52] with wavelet-based mutation operation [53] (SPSOWM), in which the simplified PSO enhances the global search capability while the wavelet-based mutation operation improves the stability and convergence of algorithm.

In SPSOWM, the velocity v has been merged into position x. The update of x is calculated by:

$$x_{id}^{k+1} = r_2 \mathrm{sign}(r_3) \cdot x_{id}^{k} + (1-r_2)c_1 r_1 \left(x_{id,\mathrm{pb}}^{k} - x_{id}^{k}\right) + (1-r_2)c_1(1-r_1)\left(x_{id,\mathrm{gb}}^{k} - x_{id}^{k}\right) \quad \text{(C.2-2)}$$

where the $\mathrm{sign}(r_3)$ means the reverse search mechanism calculated by:

$$\mathrm{sign}(r_3) = \begin{cases} -1, & r_3 < c \\ \omega, & r_3 > c \end{cases} \quad \text{(C.2-3)}$$

The r_3 is a random number as r_1 and r_2. And c is a given threshold for reverse search.

Before updating position, the particle may mutate based on Morlet wavelet function under a given probability p_{m}, in which the mutated position $mut\left(x_{id}^{k}\right)$ is calculated by:

$$mut\left(x_{id}^{k}\right)=\begin{cases}x_{id,\mathrm{gb}}^{k}+\sigma_{\mathrm{Morlet}}\times\left(px_{mxi}-x_{id,\mathrm{gb}}^{k}\right), & \sigma>0\\ x_{id,\mathrm{gb}}^{k}+\sigma_{\mathrm{Morlet}}\times\left(x_{id,\mathrm{gb}}^{k}-px_{mni}\right), & \sigma\leq 0\end{cases} \quad (C.2\text{-}4)$$

where σ_{Morlet} is the value of Morlet function depending on k and the maximum number of iterations prespecified.

By initializing the parameters and particles' position, SPSOWM firstly evaluates the fitness of position and then updates the personal and global optimums. After that, it carries out the mutation operation and then updates the position until meeting the convergence conditions.

C.3 Bayesian optimization algorithm (BO)

Bayesian optimization is mainly used to solve optimization problems where the objective function is computationally expensive or difficult to solve by traditional methods. This is done by constructing a surrogate function to fit unknown or complex objective function distributions, which makes it independent of the form of objective function. During the optimization process, it uses the acquisition function to guide the sampling of solutions, and then updates the surrogate function to continuously improve the estimation of the optimal solution.

For objective function $y=f:X\to\mathbb{R}$, in which the variables are independent of each other, initial samplings provide a set of observation history $\mathrm{H}=(x_i,f(x_i))_{i=1}^{n}$. The surrogate function based on tree-structured parzen estimator (TPE) [54] fits the conditional probability $p(x\mid y)$ of x given y, by dividing the values of objective function into two parts according to γ-quantile (corresponding to y^*), and then using two Gaussian distributions $\ell(x)$ and $g(x)$ to fit them, respectively, as following:

$$p(x\mid y)=\begin{cases}\ell(x) & \text{if } y<y^*\\ g(x) & \text{if } y\geq y^*\end{cases} \quad (C.3\text{-}1)$$

The expected improvement (EI) function, which denotes the expectation of the difference between y^* and y for a given variable x, is taken as the acquisition function and is calculated by:

$$\mathrm{EI}_{y^*}(x)=\int_{-\infty}^{y^*}(y^*-y)\,p(y\mid x)dy \quad (C.3\text{-}2)$$

Introducing Bayes rule $p(x,y)=p(y|x)p(x)=p(x|y)p(y)$, then

$$\mathrm{EI}_{y^*}(x)=\int_{-\infty}^{y^*}(y^*-y)\frac{p(x|y)p(y)}{p(x)}dy \quad (C.3\text{-}3)$$

Combining $\gamma=p(y<y^*)$ and $p(x|y)$ expression, then

$$p(x)=\int_{\mathbb{R}}p\,(x\mid y)p(y)dy=\gamma\ell(x)+(1-\gamma)g(x) \quad (C.3\text{-}4)$$

$$\int_{-\infty}^{y^*}(y^*-y)\,p(x\mid y)p(y)dy=\ell(x)\int_{-\infty}^{y^*}(y^*-y)\,p(y)dy=\gamma y^*\ell(x)-\ell(x)\int_{-\infty}^{y^*}p\,(y)dy \quad (C.3\text{-}5)$$

Therefore, the EI can be expressed as

$$EI_{y^*}(x)=\frac{\gamma y^*\ell(x)-\ell(x)\int_{-\infty}^{y^*}p(y)dy}{\gamma\ell(x)+(1-\gamma)g(x)}\propto\frac{\ell(x)}{\gamma\ell(x)+(1-\gamma)g(x)}=\left(\gamma+\frac{g(x)}{\ell(x)}(1-\gamma)\right)^{-1} \quad (C.3\text{-}6)$$

When the objective is minimization, in the next sampling step it wants $\underset{x}{\operatorname{argmax}} \mathrm{EI}_{y^*}(x)$, which is equivalent to $\underset{x}{\operatorname{argmin}} \frac{g(x)}{\ell(x)}$. This leads to points x with high probability under $\ell(x)$ and low probability under $g(x)$.

Therefore, the TPE-based Bayesian optimization can effectively use the existing sampling information to obtain better optimization results by repeating the process of "sampling-evaluation-updating probability distribution $p(x \mid y)$-sampling" until the stopping condition is satisfied.

C.4 Soft actor-critic algorithm (SAC)

Model-free deep reinforcement learning algorithms have been demonstrated in a range of challenging decision making and control tasks. However, these methods commonly suffer from two major challenges: very high sample complexity and fragile convergence properties, which require careful tuning of hyperparameters. Both challenges severely limit the applicability of these methods to complex real-world domains. The soft actor-critic (SAC) algorithm [44] is one of the best performing reinforcement learning algorithms due to its integration of actor-critic framework, off-policy strategy learning framework and maximum entropy reinforcement learning framework, which has the advantages of good stability and high sampling efficiency, and is easy to implement.

Figure C - 2 shows the interaction between agent and environment during training of RL.

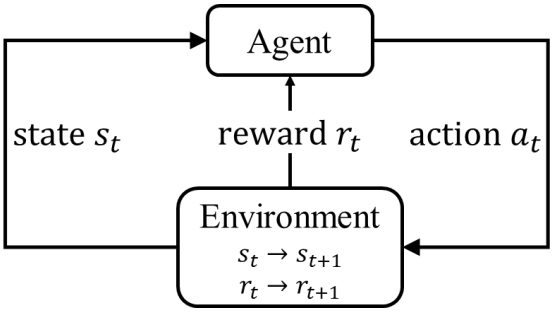


Figure C - 2 – Schematic diagram of reinforcement learning.

In RL, the reward r_t from environment depends on the environment's state $s_t \in \mathcal{S}$ and agent's action $a_t \in \mathcal{A}$ controlled by policy $\pi_\phi(s)$. The objective of agent is to maximize the reward, which is modelled as the maximization of discounted expected reward $\mathbb{E}[G_t]$:

$$\underset{\phi}{\operatorname{argmax}}\, J(\phi) = \underset{\phi}{\operatorname{argmax}}\, \mathbb{E}_{\pi_\phi}[G_t] \qquad \text{(C.4-1)}$$

Value function G presents estimation of reward r and has the form of state value function $V(s_t)$ and action value function $Q(s_t, a_t)$.

The objective function of SAC introduces an entropy term on policy π in order to maximize the reward while maintaining the diversity of actions and enhancing the stability of training, which is expressed as:

$$J(\pi) = \sum_{t=0}^{T} \mathbb{E}_{(s_t,a_t)\sim\rho_\pi} [r(s_t, a_t) + \alpha\mathcal{H}(\pi(\cdot\,|s_t))] \quad (C.4\text{-}2)$$

where $\mathcal{H}(\pi)$ represents entropy in the following form:

$$\mathcal{H}(\pi) = -\sum_{\pi_a\in\pi} \pi_a \log(\pi_a) \quad (C.4\text{-}3)$$

A larger entropy value indicates a greater randomness of the action, while a smaller entropy value indicates a stable action.

Besides, the temperature parameter α controls the ratio of exploration and exploitation. When α is larger, the entropy term becomes dominant and prefers a stochastic strategy (exploration); when α is smaller, the reward term is dominant and prefers a deterministic strategy (exploitation).

Accordingly, SAC uses the neural network model to learn the state value function V_ψ, the action value function Q_θ, and the policy π_ϕ, and updates their corresponding model parameters ψ, θ, and ϕ using stochastic gradient descent.

The soft state value function V_ψ is trained to minimize the squared residual error, as follow:

$$J_V(\psi) = \mathbb{E}_{s_t\sim D}\, MSE\left(V_\psi(s_t),\ \mathbb{E}_{a_t\sim\pi_\phi}[Q_\theta(s_t, a_t) - \log\pi_\phi(a_t|s_t)]\right) \quad (C.4\text{-}4)$$

The soft action value function Q_θ is trained to minimize the soft Bellman residual, as follow:

$$J_Q(\theta) = \mathbb{E}_{(s_t,a_t)\sim D}\, MSE\big(Q_\theta(s_t, a_t), \hat{Q}(s_t, a_t)\big) \quad (C.4\text{-}5)$$

where

$$\hat{Q}(s_t, a_t) = r(s_t, a_t) + \gamma\mathbb{E}_{s_{t+1}\sim p}[V_{\overline{\psi}}(s_{t+1})] \quad (C.4\text{-}6)$$

$$V_{\overline{\psi}} = (1-\rho)V_{\overline{\psi}} + \rho V_\psi \quad (C.4\text{-}7)$$

The policy π_ϕ is trained to minimize the expected KL-divergence, as follow:

$$J_\pi(\phi) = \mathbb{E}_{\mathbf{s}_t\sim\mathcal{D},\epsilon_t\sim\mathcal{N}}\left[\log\pi_\phi\big(f_\phi(\epsilon_t;\mathbf{s}_t)\mid\mathbf{s}_t\big) - Q_\theta\left(\mathbf{s}_t, f_\phi(\epsilon_t;\mathbf{s}_t)\right)\right] \quad (C.4\text{-}8)$$

where $f_\phi(\epsilon_t;\mathbf{s}_t)$ is the neural network model representing a_t, and ϵ_t is the input noise vector sampled from fixed distribution like spherical Gaussian.

Although the state value function V_ψ and the action value function Q_θ can represent each other, the stability of training can be improved by introducing the state value function in practice. In addition, SAC also inherits methods from Twin Delayed Deep Deterministic policy gradient algorithm (TD3) [55] such as the Double Q-learning to reduce overfitting and improve convergence. Figure C - 3 depicts the pseudo-code of the SAC training process. For detailed gradient derivation and training, please refer to [44].

Algorithm 1 Soft Actor-Critic

Initialize parameter vectors ψ, $\bar{\psi}$, θ, ϕ.
for each iteration **do**
  for each environment step **do**
    $\mathbf{a}_t \sim \pi_\phi(\mathbf{a}_t|\mathbf{s}_t)$
    $\mathbf{s}_{t+1} \sim p(\mathbf{s}_{t+1}|\mathbf{s}_t, \mathbf{a}_t)$
    $\mathcal{D} \leftarrow \mathcal{D} \cup \{(\mathbf{s}_t, \mathbf{a}_t, r(\mathbf{s}_t, \mathbf{a}_t), \mathbf{s}_{t+1})\}$
  end for
  for each gradient step **do**
    $\psi \leftarrow \psi - \lambda_V \hat{\nabla}_\psi J_V(\psi)$
    $\theta_i \leftarrow \theta_i - \lambda_Q \hat{\nabla}_{\theta_i} J_Q(\theta_i)$ for $i \in \{1, 2\}$
    $\phi \leftarrow \phi - \lambda_\pi \hat{\nabla}_\phi J_\pi(\phi)$
    $\bar{\psi} \leftarrow \tau\psi + (1 - \tau)\bar{\psi}$
  end for
end for

Figure C - 3 – Pseudo-code of the SAC [44].

References

[1] IEA. Net Zero by 2050. Paris; 2021. https://www.iea.org/reports/net-zero-by-2050.

[2] Zong Y, Wang J, You S, Su W, Ohlenschlager VF, Petersen SB, et al. Identifying the System-related Conditions and Consequences of Power-to-X Solutions for a High Renewables Penetration in Denmark. (IEEE, 2020), pp. 955-960. https://go.exlibris.link/1lSQHWtY.

[3] Crawley J, Johnson C, Calver P, Fell M. Demand response beyond the numbers: A critical reappraisal of flexibility in two United Kingdom field trials. ENERGY RES SOC SCI. 2021;75:102032. doi: 10.1016/j.erss.2021.102032.

[4] Powells G, Fell MJ. Flexibility capital and flexibility justice in smart energy systems. ENERGY RES SOC SCI. 2019;54:56-59. doi: 10.1016/j.erss.2019.03.015.

[5] Dranka GG, Ferreira P, Vaz AIF. A review of co-optimization approaches for operational and planning problems in the energy sector. APPL ENERG. 2021;304:117703. doi: 10.1016/j.apenergy.2021.117703.

[6] Department For Business EISA. Transitioning to a net zero energy system: smart systems and flexibility plan 2021. 2021. https://www.gov.uk/government/publications/transitioning-to-a-net-zero-energy-system-smart-systems-and-flexibility-plan-2021.

[7] Buckley P. Prices, information and nudges for residential electricity conservation: A meta-analysis. ECOL ECON. 2020;172:106635. doi: 10.1016/j.ecolecon.2020.106635.

[8] Zhang X, Ramírez-Mendiola JL, Li M, Guo L. Electricity consumption pattern analysis beyond traditional clustering methods: A novel self-adapting semi-supervised clustering method and application case study. APPL ENERG. 2022;308:118335. doi: 10.1016/j.apenergy.2021.118335.

[9] Davarzani S, Pisica I, Taylor GA, Munisami KJ. Residential Demand Response Strategies and Applications in Active Distribution Network Management. Renewable and Sustainable Energy Reviews. 2021;138:110567. doi: 10.1016/j.rser.2020.110567.

[10] Torriti J. Household electricity demand, the intrinsic flexibility index and UK wholesale electricity market prices. Environmental Economics and Policy Studies. 2020:1-21. doi: 10.1007/s10018-020-00296-1.

[11] Guelpa E, Verda V. Demand response and other demand side management techniques for district heating: A review. ENERGY. 2021;219:119440. doi: 10.1016/j.energy.2020.119440.
[12] Kathirgamanathan A, De Rosa M, Mangina E, Finn DP. Data-driven predictive control for unlocking building energy flexibility: A review. Renewable and Sustainable Energy Reviews. 2021;135:110120. doi: 10.1016/j.rser.2020.110120.
[13] Chehade Z, Mansilla C, Lucchese P, Hilliard S, Proost J. Review and analysis of demonstration projects on power-to-X pathways in the world. INT J HYDROGEN ENERG. 2019;44:27637-27655. doi: 10.1016/j.ijhydene.2019.08.260.
[14] Montazerinejad H, Eicker U. Recent development of heat and power generation using renewable fuels: A comprehensive review. RENEW SUST ENERG REV. 2022;165. doi: 10.1016/j.rser.2022.112578.
[15] Bansal S, Zong Y, You S, Mihet-Popa L, Xiao J. Technical and Economic Analysis of One-Stop Charging Stations for Battery and Fuel Cell EV with Renewable Energy Sources. ENERGIES. 2020;13:2855. doi: 10.3390/en13112855.
[16] Adam A, Fraga ES, Brett DJL. Options for residential building services design using fuel cell based micro-CHP and the potential for heat integration. APPL ENERG. 2015;138:685-694. doi: 10.1016/j.apenergy.2014.11.005.
[17] Di Marcoberardino G, Chiarabaglio L, Manzolini G, Campanari S. A Techno-economic comparison of micro-cogeneration systems based on polymer electrolyte membrane fuel cell for residential applications. APPL ENERG. 2019;239:692-705. doi: 10.1016/j.apenergy.2019.01.171.
[18] Arsalis A. A comprehensive review of fuel cell-based micro-combined-heat-and-power systems. Renewable and Sustainable Energy Reviews. 2019;105:391-414. doi: 10.1016/j.rser.2019.02.013.
[19] Löbberding L, Madlener R. Techno-economic analysis of micro fuel cell cogeneration and storage in Germany. APPL ENERG. 2019;235:1603-1613. doi: 10.1016/j.apenergy.2018.11.023.
[20] Bornapour M, Hooshmand R, Khodabakhshian A, Parastegari M. Optimal stochastic scheduling of CHP-PEMFC, WT, PV units and hydrogen storage in reconfigurable micro grids considering reliability enhancement. ENERG CONVERS MANAGE. 2017;150:725-741. doi: 10.1016/j.enconman.2017.08.041.
[21] Dinh HN, Ishihara T. Distributed peer-to-peer energy trading for residential fuel cell combined heat and power systems. INT J ELEC POWER. 2021;125. doi: 10.1016/j.ijepes.2020.106533.
[22] Arsalis A. A comprehensive review of fuel cell-based micro-combined-heat-and-power systems. Renewable and Sustainable Energy Reviews. 2019;105:391-414. doi: 10.1016/j.rser.2019.02.013.
[23] Danish Energy Agency. Technology Data for Energy Storage. 2020. https://ens.dk/sites/ens.dk/files/Analyser/technology_data_catalogue_for_energy_storage.pdf.
[24] Danish Energy Agency. Technology Data for Renewable Fuels. 2022. https://ens.dk/sites/ens.dk/files/Analyser/technology_data_for_renewable_fuels.pdf.
[25] Deng K, Liu Y, Hai D, Peng H, Löwenstein L, Pischinger S, et al. Deep reinforcement learning based energy management strategy of fuel cell hybrid railway vehicles considering fuel cell aging. ENERG CONVERS MANAGE. 2022;251:115030. doi: 10.1016/j.enconman.2021.115030.
[26] Barbir F. PEM Fuel Cells: Theory and Practice. Elsevier/Academic Press. 2012. ISBN: 978-0-12-387710-9.
[27] Danish Energy Agency. Technology Data for Individual Heating Plants. 2021. https://ens.dk/sites/ens.dk/files/Analyser/technology_data_catalogue_for_individual_heating_installations.pdf.
[28] Battelle Memorial Institute. Manufacturing Cost Analysis of 100 and 250 kW Fuel Cell Systems for Primary Power and Combined Heat and Power Applications. 2017. https://www.energy.gov/sites/default/files/2016/07/f33/fcto_battelle_mfg_cost_analysis_pp_chp_fc_systems.pdf.
[29] Danish Energy Agency. Technology Data for Generation of Electricity and District Heating. 2022. https://ens.dk/sites/ens.dk/files/Analyser/technology_data_catalogue_for_el_and_dh.pdf.
[30] Low Carbon London project: Data from the dynamic time-of-use electricity pricing trial. 2013. https://innovation.ukpowernetworks.co.uk/projects/low-carbon-london/.
[31] Schofield JR. Dynamic time-of-use electricity pricing for residential demand response: Design and analysis of the Low Carbon London smart-metering trial: Imperial College London; 2015.
[32] Networks UP. Impact of Electric Vehicles and Heat Pump Loads on Network Demand Profiles. 2014. https://innovation.ukpowernetworks.co.uk/wp-content/uploads/2021/04/LCL-DNO-Report-B2-Impact-of-Electric-Vehicle-and-Heat-Pump-loads-on-Network-demand-profiles.pdf.
[33] Miller C, Picchetti B, Fu C, Pantelic J. Limitations of machine learning for building energy prediction: ASHRAE Great Energy Predictor III Kaggle competition error analysis. Science & technology for the built environment. 2022;28:610-627. doi: 10.1080/23744731.2022.2067466.
[34] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews. 2018;81:1192-1205. doi: 10.1016/j.rser.2017.04.095.
[35] Sevlian R, Rajagopal R. A scaling law for short term load forecasting on varying levels of aggregation. INT J ELEC POWER. 2018;98:350-361. doi: 10.1016/j.ijepes.2017.10.032.
[36] Hong T, Fan S. Probabilistic electric load forecasting: A tutorial review. INT J FORECASTING. 2016;32:914-938. doi:

10.1016/j.ijforecast.2015.11.011.
[37] Koenker R, Bassett G. Regression Quantiles. ECONOMETRICA. 1978;46:33-50. doi: 10.2307/1913643.
[38] Hochreiter S, Schmidhuber J. Long Short-Term Memory. NEURAL COMPUT. 1997;9:1735-1780. doi: 10.1162/neco.1997.9.8.1735.
[39] Ahmad T, Zhang H, Yan B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. SUSTAIN CITIES SOC. 2020;55:102052. doi: 10.1016/j.scs.2020.102052.
[40] Xiaohai Zhang ML. gensbo: a GENeral Simulation Based Optimizer. PyPI; 2021. https://pypi.org/project/gensbo/.
[41] Bergstra J, Yamins D, Cox D. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. Proceedings of the 30th International Conference on Machine Learning, PMLR, (2013), pp. 115-123. http://proceedings.mlr.press/v28/bergstra13.html.
[42] 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;10:27853-27862. doi: 10.1109/ACCESS.2022.3156581.
[43] Vázquez-Canteli JR, Nagy Z. Reinforcement learning for demand response: A review of algorithms and modeling techniques. APPL ENERG. 2019;235:1072-1089. doi: 10.1016/j.apenergy.2018.11.002.
[44] Haarnoja T, Zhou A, Abbeel P, Levine S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research; 2018. https://proceedings.mlr.press/v80/haarnoja18b.html.
[45] Haarnoja T, Zhou A, Hartikainen K, Tucker G, Ha S, Tan J, et al. Soft Actor-Critic Algorithms and Applications. ArXiv. 2018. doi: 10.48550/arXiv.1812.05905.
[46] Weng J, Chen H, Yan D, You K, Duburcq A, Zhang M, et al. Tianshou: a Highly Modularized Deep Reinforcement Learning Library. JMLR, 2022. http://jmlr.org/papers/v23/21-1127.html.
[47] Jin Y, Wang H, Chugh T, Guo D, Miettinen K. Data-Driven Evolutionary Optimization: An Overview and Case Studies. IEEE T EVOLUT COMPUT. 2018;PP:1. doi: 10.1109/TEVC.2018.2869001.
[48] Winkler RL. A Decision-Theoretic Approach to Interval Estimation. J AM STAT ASSOC. 1972;67:187-191. doi: 10.1080/01621459.1972.10481224.
[49] Aston Zhang ZCLM. Dive into Deep Learning. arXiv preprint; 2021. https://doi.org/10.48550/arXiv.2106.11342.
[50] James Kennedy RE. Particle Swarm Optimization. International Conference on Neural Networks, 1995.
[51] Wang D, Tan D, Liu L. Particle swarm optimization algorithm: an overview. SOFT COMPUT. 2018;22:387-408. doi: 10.1007/s00500-016-2474-6.
[52] Clerc M. The swarm and the queen: Towards a deterministic and adaptive particle swarm optimization. 1999. ISBN: 0-7803-5536-9.
[53] Tian Y, Gao D, Li X. Improved Particle Swarm Optimization with Wavelet-Based Mutation Operation. Tan Y, Shi Y, Ji Z, eds. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012), pp. 116-124. https://doi.org/10.1007/978-3-642-30976-2_14.
[54] Bergstra J, Bardenet REM, Bengio Y, K E Gl BAZ. Algorithms for Hyper-Parameter Optimization. NIPS'11, Red Hook, NY, USA; 2011.
[55] Fujimoto S, Hoof H, Meger D. Addressing Function Approximation Error in Actor-Critic Methods. Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research; 2018. https://proceedings.mlr.press/v80/fujimoto18a.html.