Energy Conversion and Management: X 15 (2022) 100264

Contents lists available at ScienceDirect

Energy Conversion and Management: X

journal homepage: www.sciencedirect.com/journal/energy-conversion-and-management-x

# Advanced control techniques for CHP-DH systems: A critical comparison of Model Predictive Control and Reinforcement Learning

A. Mugnini [a,*], F. Ferracuti [a], M. Lorenzetti [b], G. Comodi [a], A. Arteconi [a,c]

[a] Dipartimento di Ingegneria Industriale e Scienze Matematiche, Università Politecnica delle Marche, Via Brecce Bianche 12, 60131 Ancona, Italy
[b] Astea S.p.A, 60027 Ancona, Italy
[c] Department of Mechanical Engineering, KU Leuven, B-3000 Leuven, Belgium

ARTICLE INFO

*Keywords:*
Model Predictive Control
Reinforcement Learning
Combined Heat and Power
District Heating
Operational Data

ABSTRACT

District heating (DH) network is a key infrastructure to decarbonize the heating sector through the centralized production of heat distributed to final users. The implementation of advanced control techniques is increasingly common in the field of energy optimization since they can provide a more efficient way of minimizing energy demand by appropriate scheduling of the control variables.

The aim of this work is to present the application of two control strategies, i.e., Model Predictive Control (MPC) and Reinforcement Learning (RL), to a system based on a DH network supplied by a Combined Heat and Power plant (CHP-DH plant). The analyzed case study is a real CHP-DH plant operating in the small Italian town of Osimo (central Italy). The DH network currently connects more than 1200 users, generating peak heat demand of about 9.7 $MW_{th}$. The heat generator is composed of a natural gas fueled internal combustion engine coupled with natural gas boilers.

The work provides a comparison between the current control strategy (deduced from measured data) and the performance of the CHP-DH plant controlled with an MPC and an RL control. The results showed the effectiveness of the two controls in satisfying the thermal demand of the users, while minimizing the thermal losses towards the ground. Both MPC and RL allow to implement control strategies different from the current control in terms of supply temperature and flow rate circulating in the network. Referring to the winter months, in which the current operation of the system tends to prefer high supply temperatures, the advanced controls made it possible to reduce the thermal heat supply by reducing the thermal losses of about 3.9 % with the MPC and 6.54 % with the RL, corresponding to emission avoidances up to 23.3 $tCO_2$ and 12.6 $tCO_2$, respectively.

The paper, as well as showing the application of the controls, contains a critical discussion of all the positive aspects and weaknesses found in the application of the MPC and the RL control to the case study.

## 1. Introduction

In recent years, global energy policies are increasingly aimed at accelerating the transition to a fully decarbonized energy system. Energy efficiency, circular economy and reduction of greenhouse gas emissions are the main objectives that the European Union has set for the medium-long term [1]. Several technological solutions for the different sectors responsible for most of the energy consumption (i.e., industrial, buildings, transport, …) have been identified and encouraged. One of the most promising areas to address is heat demand. According to the International Energy Agency, heat is the largest energy end-use [2]. Considering both the residential and industrial sectors, the energy demand for heating represents about 50 % of the global energy use and is responsible for more than 40 % (13.1 Gt in 2020) of $CO_2$ emissions [2].

The European Union has basically identified two ways to achieve the energy transition of the thermal sector, namely (i) electrification of heating demand and (ii) efficient heat and power production via Combined Heat and Power (CHP) plants and District Heating (DH) networks [3]. In particular, DH can play a very important role in decarbonising the heating sector, thus the European Union has planned strong investments to increase its spread in urban areas by 2050 [4].

There are many advantages introduced by a wide spread of DH. First, the replacement of localized production by larger distributed generation plants. This allows to satisfy the demand with an increased efficiency of

* Corresponding author.
*E-mail address:* a.mugnini@univpm.it (A. Mugnini).

https://doi.org/10.1016/j.ecmx.2022.100264
Received 12 May 2022; Received in revised form 25 June 2022; Accepted 1 July 2022
Available online 5 July 2022

**Nomenclature**

| | |
|---|---|
| $\pi$ | Policy for RL |
| **A** | Matrix for state space model |
| a | Action for RL |
| A | Set of action for RL |
| AC | Actor-critic agent for RL |
| **B** | Matrix for state space model |
| C | Thermal capacity (J $K^{-1}$) |
| c | Specific heat (J $kg^{-1}$ $K^{-1}$) |
| CHP | Combined Heat and Power |
| ct | Control (1/0) |
| DH | District Heating |
| f | Factor (–) |
| G | Sum of future discounted rewards for RL |
| HVAC | Heating, Ventilation and Air-Conditioning |
| i | Index |
| ICE | Internal Combustion Engine |
| j | Index |
| K | Themal conductance (W $K^{-1}$) |
| k | Discrete time (s) |
| $\dot{m}$ | Flow rate (kg $s^{-1}$) |
| MDP | Markov Decision Process |
| MPC | Model Predictive Control |
| N | Horizon (s) |
| o | Observation for RL |
| P | Terminal weight |
| PID | Proportional-Integral-Derivative |
| POMDP | Partially Observable Markov Decision Process |
| Q | Weight for predicted output error |
| q | Action-value function for RL |
| $\dot{Q}$ | Thermal power ($W_{th}$) |
| r | Reference output |
| R | Weight for input rate error |
| RL | Reinforcement Learning |
| RMSE | Root Mean Squared Error |
| Rw | Reward for RL |
| s | State for RL |
| S | Set of states for RL |
| T | Temperature (°C) |
| t | Time (s) |
| TES | Thermal Energy Storage |
| u | Input |
| x | State model |
| y | Output |
| $\gamma$ | Discount factor |
| $\Delta u$ | Predicted input rate |
| $\nu$ | Critic for RL |

*Subscripts*

| | |
|---|---|
| DH | District Heating |
| B | Baseload |
| c | Corrective |
| e | External |
| L | Losses |
| max | Maximum value |
| min | Minimum value |
| p | Predicted (referred to predictive horizon) |
| r | Return |
| r,sp | Setpoint for the return temperature |
| s | Supply |
| SH | Space heating |
| t | Thermostat |
| U | Users connected |
| u | Control (referred to control horizon) |
| w | Water |

energy conversion [5]. Moreover, DH allows to efficiently integrate different energy sources, including waste energy and renewable sources. In this regard, Yuan et al. [6] proposed an interesting case study in which the heating demand of a community in Aalborg (Denmark), estimated to reach 1.56 $TWh_{th}$ per year, is satisfied in a completely clean way with a DH system. The heat supply, in fact, comes both from the exploitation of excess heat from an industrial user, and from the use of renewable sources which supply heat pumps.

The scenario assessed by the Yuan et al. is an example to understand the great potential of DH to meet the thermal demand of a network of users in a sustainable way. This ability, however, is not only due to the possibility of integrating clean energy sources but also to the possibility of managing them in a flexible way. In fact, as many studies show [7], DH networks can also allow to decouple demand from generation in different ways. The heat transfer fluid contained in the pipes can be used as a thermal storage medium. In addition, the network can integrate several added Thermal Energy Storage (TES) devices to increase thermal inertia [8]. In this sense, Zhang et al [9] proposed an interesting mapping of the potential of the different TESs that can be used in low temperature DH networks. From the results of their simulations, the authors concluded that the best flexibility performance can be achieved with a centralized storage tank while discouraging the use of building envelope mass and inertia of network.

To activate the flexibility in DH networks an effective means is represented by advanced control techniques [8] and, in particular, optimal controls for this purpose are quite widespread. In the literature several studies can be found in which the ability of optimal controls to activate flexibility in DH networks is assessed. For instance, Laakkonen et al. [10] implemented an optimized control to minimize pumping costs and heat losses by exploiting the flexibility provided by the thermal inertia of the DH pipes. Predicting the time delays of DH response, the control proposed by Laaksonen et al. is based on determining the supply temperatures to be set. Also, Bavière and Vallée [11] proposed a similar application. Indeed, they implemented a controller able to optimize the heat distribution through a suitable programming of the supply temperatures and differential pressure at the production level. Their aim was to minimize economic expenditure by flexibly exploiting network management through the prediction of the dynamic behaviour of the system.

Although these are just some of the examples of the application of optimal controls to DH networks, in this work we decided to focus on a specific type of optimal control (Model Predictive Control, MPC), considering its increasing diffusion for the advanced management of energy systems [12]. An MPC is a constrained optimal control that calculates the control actions by minimizing a given objective function over a finite prediction horizon [13]. The MPC application to manage DH networks is not new. Currently, there are many studies in literature proving the effectiveness of MPC for DH networks. For instance, Verrilli et al. [14] implemented a MPC to activate energy flexibility in a DH system with a TES and flexible loads. The authors considered a network of almost 600 users in Ylivieska (Finland) and demonstrated that the management through MPC of the heating power plant (consisting of CHP, grate boiler, or oil boiler) and the TES allows to estimate a potential cost reduction of about 7.5 % compared to a scenario without MPC. Another example of applying MPC to a DH network is the work proposed by Hering et al. [15]. In this case the authors studied a low temperature DH system powered by heat pumps. Their MPC was formalized to optimize the operation of heat pumps in combination with

thermal energy storages. Controlling the flow rate and the return temperature, Hering et al. obtained savings of electrical energy consumption up to 5.49 %. Also, Saloux and Candanedo [16] demonstrated the effectiveness of MPC for DH system. They implemented a MPC aimed to minimize the primary energy use in a solar DH with dedicated energy storage. The case study considered is the Drake Landing Solar Community (52 homes in Okotoks, Canada) which is composed of solar thermal collectors along with a thermal energy storage and a back-up natural gas boiler. With the MPC acting on variable speed circulation pumps, the authors estimated pumping power savings of 47 %, energy costs reduction of 38 % and a 32 % decrease of the greenhouse gas emissions. Another interesting example of application of an MPC to a thermal network is the one proposed by Vivian et al. [17]. Indeed, they formulized a MPC to optimize a centralised CHP and storage unit with the aim of reducing the overall operational costs in a fifth-generation thermal network. With their study, Vivian et al. also demonstrated the potential effectiveness of MPC and obtained an estimated saving of around 11 % in comparison to a traditional rule-based control.

These are just some of the several studies in literature that evaluate the application of MPCs to thermal networks and/or their generation systems. In recent years, however, interest in assessing the application of fully data-based controls for energy systems is also increasing. In particular, Machine Learning (ML) techniques look very promising [18]. ML algorithms are based purely on data and enable models to learn by themselves once the learning algorithm is determined [19]. Reinforcement Learning (RL) is a particular type of ML. RL models learn to perform serial actions according to situations to maximize the reward signal by trial-and-error search [20]. Applications of RL algorithms are increasingly widespread in buildings [21]. Several applications concern the control of building heating and/or cooling systems to increase the efficiency [22] or to unlock the flexibility (e.g, Demand Response strategies) [23]. Furthermore, RL is also chosen as a control technique when dealing with multiple energy carriers [24].

Considering the thermal networks, ML and RL algorithms are currently mainly used to predict the thermal load of the network. Indeed, Sakkas and Abang [25] applied a data-driven ML based on artificial neural networks to forecast the thermal load of the DH network of Cottbus (Germany). Also, Wei et al. [26] conducted a similar study for a residential DH system. In particular, the authors compared different ML algorithms (e.g., Support Vector Regression, XGBoost and Long Short-Term Memory neural network) to predict the thermal load of a DH network in Shanghai (China). The same objective is also posed by the work of Geysen et al [27], who implemented different algorithms of ML (i.e., Linear Regression Artificial Neural Networks, Support Vector Machines and extremely randomized extra tree regressors) to predict the thermal demand of a DH network based on outdoor temperature forecast, historic thermal load information and historic control signals. Maljkovic and Basic [28], indeed, evaluated the main parameters that influence the heat demand in DH networks by applying ML techniques. The authors implemented several algorithms (i.e., Regression Trees, Random Forest and Regression Support Vector Machines) to predict the heat demand of a DH network based on actual billing data for 260 buildings and data for a specific demonstration building in Zagreb (Croatia). The interesting aspect of their work is that they have demonstrated that the most influential parameter on the heat consumption of a single final consumer is the overall consumption of the building, while the second is users' behavior.

It is important to note that in all the above-mentioned works, the ML algorithm is not used to establish control logics for the network and/or its generator. A case in which an RL algorithm is integrated within the control is the one presented by Solinas et al. [29]. Indeed, Solinas et al. proposed an interesting application of RL to a DH network to produce peak shaving events. To limit the peak in a DH network located in Turin (Italy), the authors proposed a combined control: a thermodynamic model was used to assess the response of the buildings to energy profile modifications and an agent-based model to represent the end-users' adaptability to imposed temperatures variations. Then, the RL algorithm was formulated to choose the best control action (a set of anticipations and delays to energy profiles) to reduce peak thermal demand and limit the dissatisfaction of the most sensitive users.

These are some of the examples that are available in the literature involving the use of ML or RL algorithms on thermal networks. However, applications of RL algorithms for direct management of DH network and/or its generation system are not very widespread. The literature analysis allows to observe that the studies on MPC controls applied to the management of DH networks are quite numerous and all show a good performance of the control in achieving the target objective. On the other hand, the use of RL techniques to control DH is less common. However, in other applications involving energy systems, the discussion about choosing an optimized control based on system model (e.g., MPC) rather than a completely data-based approach (e.g., RL) for energy systems is still open. Some papers have addressed this issue for some case studies. For instance, Brandi et al. [30] compared an online and offline deep RL with MPC for thermal energy management in an office building and Ceusters et al. [31] compared MPC and RL for a high-level control in multi-energy systems. Brandi et al. [30] suggested that the deep RL agent trained online may represent a promising solution to overcome the barrier represented by the modelling requirements of MPC and offline-trained deep RL approaches. Even Ceusters et al. [31] concluded that RL can be an adequate control technique for multi energy systems that can also outperform MPC, given a sufficient training in terms of time and memory. However, they also highlighted how the performance of the RL greatly depends on the selection of adequate a priori unknown hyper-parameters.

To the best of the authors' knowledge, a comparison between the application of MPC and RL for an energy system composed of a DH network and its heat generation systems is not yet available in literature. Although, as emerged from the literature analysis, the application of an MPC control for the management of a thermal network is quite widespread, in this paper we propose an evaluation of its potential application to a real case study in comparison with the application of an RL-based control technique, which, as mentioned, is less common. The real case study is a CHP-DH plant, operating in central Italy (Osimo), for which measured data are available. The aim of the work is twofold. Firstly, to compare in terms of performance and effectiveness the application of the two control techniques (MPC and RL) for a real CHP-DH plant. Secondly, provide a critical discussion of the strengths and limitations that have emerged during the modelling and training of the controls. This paper wants to propose some food for thought on the application of these two controls to a system similar to the one considered.

The paper is structured as follows: Section 2 describes the methods and the considered case study. Section 3 discusses the results obtained, also reporting a critical analysis on the application of the two control techniques. Finally, Section 4 summarizes the main conclusions obtained.

## 2. Materials and methods

In this Section, the modelling techniques of the MPC and the RL controls applied to the case study will be described. The Section is divided into three subsections: the first (subsection 2.1) contains the description of the case study; in subsection 2.2 the formulation of the MPC is described in detail, while in the last subsection (2.3) the RL is presented.

### 2.1. Description of the case study

The case study consists of a CHP plant which feeds a DH network in central Italy. The CHP-DH plant has also been presented in detail in other studies ([5 32 33]), however this short Section reports the main features useful for understanding the setting and modelling of the

controls listed below. In the first subsection (2.1.1) the DH network is described. In subsection 2.1.2 the main features of the heat generator are presented, while in the last (subsection 2.1.3) the control strategy currently used is reported.

### 2.1.1. District Heating network

As mentioned, the DH network is located in Osimo, a small town in the centre of Italy (43°29′09.89″N, 13°28′55.56″E). The network is the only one present in the Marche region and connects about 1278 utilities, of which 94 % are residential buildings and the remaining 6 % are public/commercial customers. In terms of heat demand, residential users are responsible for 49 % of demand, while the remaining 51 % is absorbed by public/commercial users. The DH network is 45 km long and contains approximately 444 $m^3$ of water, used as a heat transfer fluid. The pipes are made of steel and insulated with polyurethane. The thermal power transported by the network reaches a peak around 9.7 $MW_{th}$, in winter while falls below 1 $MW_{th}$ in summer. Fig. 1 represents the duration curve of the thermal demand met by the DH network for the year 2018 (measured data).

The network has an incidence of thermal losses towards the ground which varies considerably with the seasons. It was estimated that for the entire year 2018, the thermal losses amounted to approximately 5983 $MWh_{th}$, about 29 % of total heat demand [32]. As shown in Fig. 2, the percentage share of losses increases from the winter to the summer months, in which they reach peak values of 63.5 %.

### 2.1.2. Thermal power plant

The DH network described above is supplied by a thermal plant consisting of a CHP unit and 3 natural gas boilers. The main generator of the CHP is a Natural Gas (NG) fuelled Internal Combustion Engine (ICE), with a rated thermal power of 1.3 $MW_{th}$ and 1.2 $MW_{el}$ of rated electricity production. Due to its small size, the CHP engine only covers the base-load from the DH (Fig. 1). The remaining part is supplied by the integration boilers (each of 4.6 $MW_{th}$,). Table 1 summarises the main size characteristics of the generating system.

Fig. 3 shows a schematic of the coupling of the generation plant with the DH. It can be noted that the CHP engine is connected in series with the two boilers, which are in parallel with each other. A plate heat exchanger allows the transfer of thermal power from the generation side to the DH network.

### 2.1.3. Current control strategy

The present control strategy can be subdivided into two levels: (i) the high-level and the (ii) the low-level control. The high-level control concerns the choice of the generation technologies involved and their order of operation, while the low-level control sets the supply temperatures and the water flow rates to the DH. Currently, the operation strategy of the plant changes during the months of the year and aims to maximize the hours of operation of the CHP, avoiding that it works at too low load modulations. It is considered that below 60 % (thermal demand below 780 $kW_{th}$) the performance is too poor, and the engine is shutdown. Therefore, in the winter season (November–March), in which there is always a demand above the minimum modulation threshold, the CHP unit is always operating. In these months, the boilers are also switched on to cover the excess of the demand. The first one is activated, then if the demand cannot be satisfied, the second one also enters in

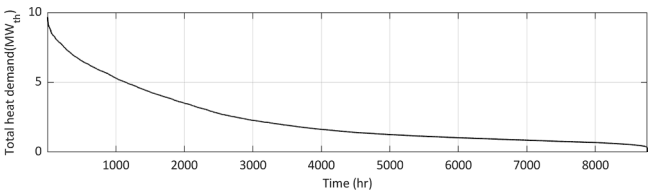


**Fig. 1.** Duration curve for the DH heat load (year 2018).

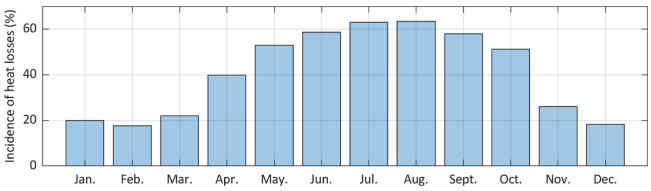


**Fig. 2.** Monthly percentage impact of thermal losses on demand (year 2018).

**Table 1**
Technical data of the thermal power plant.

| Data | Value |
|---|---|
| CHP engine electric power ($MW_{el}$) | 1.2 |
| CHP engine thermal power ($MW_{th}$) | 1.3 |
| Single operating boiler thermal power ($MW_{th}$) | 4.6 |
| Thermal CHP efficiency (%) | 42 |
| Electrical CHP efficiency (%) | 41 |
| Thermal Boiler efficiency (%) | 96.2 |

operation. In the mid-season (April–mid-June and mid-September–October), the CHP is switched off during the night (from 8.00 pm to 7.00 am), when only boilers are activated. During the hours of the day the operation is the same as in the winter months. In the summer season (mid-June–mid-September), the CHP does not work, and the thermal power is entirely produced by the boilers.

Even the regulation of the water supply temperature follows a seasonal schedule. In general, the aim is to keep it as low as possible to limit losses, compatibly with health constraints (avoid legionella) and users' demand requirement. Currently the following seasonal values are chosen with a rule of thumb method: 95 °C in the winter season (November–March), from 78 °C to 85 °C in the mid-season (April–mid-June and mid-September–October) and 75 °C in summer (mid-June–mid-September). Currently a PID (Proportional-Integral-Derivative) controller is used as low-level control. It is calibrated to maintain an empirically assumed water supply temperature value and the desired pressure level in the circuit. In this way the system is regulated in flow rate as the return temperature value is also imposed close to a set set-point value (around 60–63 °C). The limit of flow rate in the circuit is determined by the characteristics of the hydraulic pumps. The maximum volumetric flow rate is 320 $m^3\ h^{-1}$, but in the current operation generally 250 $m^3\ h^{-1}$ is not exceeded.

Table 2 summarizes the main characteristics of the current control strategy of the CHP-DH plant distinguishing the season and high-level from low-level control.

## 2.2. Model Predictive Control

Given its ability to merge principles of feedback control and numerical optimization, MPC is one of the most used controls to optimally manage the energy demand in buildings [34]. MPC is an advanced control technique that selects the control actions based on a dynamic model of the system and the resolution of an optimization problem. The system model shall be capable of capturing dynamic of the system when it is subjected to external inputs which can be controlled (or manipulated variables) or uncontrolled (or external disturbances) [35]. Then the optimization problem has the task of identifying the best sequence of control actions to achieve a specific objective, forward in time. In other words, MPC follows a "receding horizon" logic: the optimization problem is solved at each timestep moving forward the prediction horizon and the optimal control sequence is updated by applying to each time interval only the first value of the manipulated variables [34].

As mentioned, this work shows the application of an MPC strategy on a system consisting of a DH network supplied by a CHP plant. Therefore, the model of the system must be able to represent the thermal dynamics

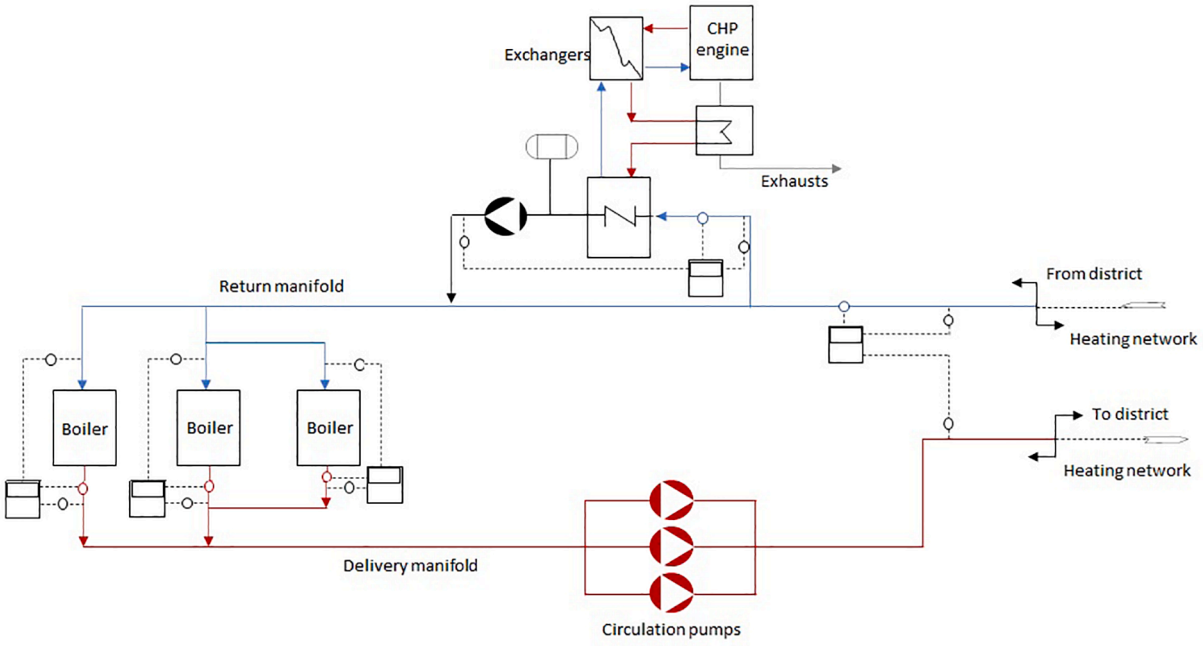


**Fig. 3.** Scheme of the thermal plant connected to the DH network.

**Table 2**
Main characteristics of the CHP-DH plant current control strategy.

| Season | High-level control | Low-level control |
|---|---|---|
| Winter (November–March) | • CHP switched on 24 hours a day<br>• Boilers integrate CHP | • Variable flow rate<br>• Return temperature in the 60-63 ° C range<br>• Supply temperature of 95 ° C |
| Mid-season (April–mid-June and mid-September–October) | • CHP switched off from 8.00 pm to 7.00 am<br>• From 8.00 pm to 7.00 am only boilers are activated. In the remaining hours boilers integrate CHP | • Variable flow rate<br>• Return temperature in the 60-63 ° C range<br>• Supply temperature in the 78 -85 °C range |
| Summer (mid-June–mid-September) | • CHP switched off<br>• Only boilers activated | • Variable flow rate<br>• Return temperature in the 60-63 ° C range<br>• Supply temperature of 75 ° C |

of the DH network. Regarding the objective of the control, the MPC has to minimize the overall energy consumption to guarantee the user's thermal demand satisfaction. The modelled MPC replaces the current low-level control (subsection 2.1.3). This means that, the control actions are addressed to the output variables of the heat generator (boilers and CHP engine): the water supply temperature and the flow rate.

This Section is divided as follows: in the first subsection (2.2.1) the model of the DH network is described. In subsection 2.2.2 the optimization problem is defined and, finally, the last subsection (2.2.3) contains the description of the control settings and constraints.

### 2.2.1. *Dynamic model*

To represent the dynamic behaviour of the DH network a non-linear grey box model has been developed. The formulation of the model is represented by Equation (1). The formulation is inspired by a previous white box version for the same DH network [33].

$$C_{DH}\frac{dT_r}{dt} = \dot{m}_{DH}c_w(T_s - T_r) - \dot{Q}_U - \dot{Q}_L + f_c(T_r - T_{r,sp}) \quad (1)$$

Equation (1) allows to evaluate the temporal evolution of the water return temperature ($T_r$ in [°C]), considering both the thermal inertia of the network, represented by the a thermal capacity ($C_{DH}$ in [J $K^{-1}$]), the power associated with the flow (i.e., the water flow rate, $\dot{m}_{DH}$, which multiplies the temperature difference between the supply and the return temperature, $T_s$, and $T_r$), the thermal demand required by the connected users ($\dot{Q}_U$ in [W]) and the thermal losses towards the ground ($\dot{Q}_L$ in [W]). Furthermore, the last term, containing the corrective factor ($f_c$ in [W $K^{-1}$]) represents the limit to the variation with respect to an imposed set point temperature ($T_{r,sp}$ in [°C]). The thermal demand ($\dot{Q}_U$) is evaluated as the sum of two contributions: the demand for space heating and a base load ($\dot{Q}_B$ in [W]), as represented by Equation (2).

$$\dot{Q}_U = K_{SH}ct_{SH}(T_t - T_e) + \dot{Q}_B \quad (2)$$

The demand for space heating is represented by the multiplication of a thermal loss factor $K_{SH}$ in [W $K^{-1}$], for a control variable that represents the presence of heat demand of users ($c_{SH}$, 1/0 signal: 1 heating systems on / 0 heating systems off) and for the temperature difference between the set point imposed by the users' thermostat ($T_t$ in [°C]) and the external air temperature ($T_e$ in [°C]). The thermal losses are instead modelled with a loss coefficient $K_L$ (in [W $K^{-1}$]) which multiplies the difference in temperature between the temperature of the water inside the pipes (obtained as the average between the water supply and the return temperature) and the ground temperature (assumed equal to the temperature of the external air).

With this formulation, the model can be written as a state space model in which the state (the return temperature, $T_r$) also coincides with the output of the model itself. Instead, the inputs are the supply temperature ($T_s$), the water flow rate ($\dot{m}_{DH}$), the external temperature ($T_e$), the set-point profile, both for the air thermostat ($T_t$) and the return temperature ($T_{r,sp}$) together with the power request signal for space heating ($ct_{SH}$).

Those described so far are the physical meanings of the quantities involved in the model. However, given the availability of measured data of the real CHP-DH plant (year 2018), the numerical values of the parameters ($C_{DH}, K_{SH}, \dot{Q}_B, K_L, f_c$) were identified through a model training

process. The details on the training result will be shown in subsection 2.2.3.

### 2.2.2. Optimization problem

The objective of the MPC is to minimize the overall heat demand covered by the heat generator, with the aim of reducing the overall fuel consumption (natural gas for both the boilers and the CHP engine) and keeping the return temperature within the acceptable range of variation. Equation (3) represents the MPC problem formulation.

$$\min_{\Delta u} \sum_{i=0}^{N_P-1} \|Q(y_{k+i|k} - r(k))\|_2^2 + \sum_{j=0}^{N_u-1} \|R\Delta u_{k+j|k}\|_2^2 + \|P(y_{k+N_p|k} - r(k))\|_2^2 \quad (3)$$

Where $y_{k+i|k}$ represent the output predicted $i$ steps ahead, $r(k)$ is the reference output over the prediction/control horizon, $\Delta u_{k+j|k}$ and $u_{k+j|k}$ are respectively the predicted input rate and input magnitude, $N_p$ is the prediction horizon while $N_u$ the control horizon. The matrices $Q$, $R$ and $P$ are used for weighting respectively the predicted output error, input increments and final output value (terminal weight).

Referring to the case study, the predicted outputs ($y_{k+i|k}$) coincide with the total heat demand $\dot{Q}$ (in [W]) and the return temperature ($T_r$). The first is calculated simply as:

$$\dot{Q} = \dot{m}_{DH} c_w (T_s - T_r) \quad (4)$$

where $c_w$ (in [J kg$^{-1}$ K$^{-1}$]) is the specific heat of the water. The second output $T_r$ is the only one controlled. In fact, the reference output ($r(k)$) is assumed to be constant over the prediction/control horizon (return temperature set-point, $T_{r,sp}$). The control actions (i.e., controlled inputs) act on the output variables of the heat generator, that are, the water flow rate ($\dot{m}_{DH}$) and the water supply temperature ($T_s$).

Summarizing therefore, the optimization problem within the MPC sets, at each timestep (15 min), the values of $\dot{m}_{DH}$ and $T_s$ to minimize the overall thermal demand that the generator must cover ($\dot{Q}$), while maintaining the return temperature close to the imposed set point value. The objective pursued by the MPC can also be seen as the minimization of the thermal losses towards the ground, which are the term that significantly depends on the supply temperature.

The constraints to the optimization problem are described by the following Equations:

$$x_{k|k} = x(k) \quad (5)$$

$$x_{k+i+1|k} = \mathbf{A}x_{k+i|k} + \mathbf{B}u_{k+i|k} \quad (6)$$

$$\Delta u_{k+j|k} = 0 for j \geq N_u \quad (7)$$

$$\Delta u_{min} \leq \Delta u_{k+j|k} \leq \Delta u_{max} \quad (8)$$

$$y_{min} \leq y_{k+i|k} \leq y_{max} \quad (9)$$

in which, Equations (5) and (6) contain the model of the DH network in state space form, Equations (7) and (8) represent the boundary conditions imposed on the control input (i.e., supply temperature and water flow rate) and Equation (9) establishes the boundary conditions imposed on the predicted outputs (i.e., the return temperature).

To sum up, Fig. 4 schematises the structure of the modelled MPC control.

### 2.2.3. Parameters and settings

In the previous two subsections, both the system model and the formulation of the MPC have been theoretically presented. In this subsection, it is described how these two aspects are translated in the analysed CHP-DH plant. First the training and testing results for the DH model are presented. Secondly the numerical details of the settings related to the MPC optimization problem are given. Table 3 shows the numerical values of the grey box model parameters obtained from the training. To highlight the physical meaning of the parameters, Table 3 also shows the range of variation admitted for each parameter assumed by the physical knowledge of the system [33].

The performance of the training and validation of the model are instead contained in Table 4. In particular, the Root Mean Squared Error (RMSE) values related to the return temperature ($T_r$) and to the overall heat demand ($\dot{Q}$) are presented. The values were calculated for both the training data set (data for the first 3 weeks of January 2018) and for two different testing periods: a shorter period covering two winter and mid-season weeks (22 January/4 February 2018 and 22 March/4 April 2018) and the entire available data set (year 2018). As can be seen, the RMSE values obtained for $T_r$ in the testing data set does not differ much from the value obtained in training: it remains between 1.59 °C and 1.76° C. The model also performs well in predicting the overall heat demand. In fact, if the RMSEs obtained for $\dot{Q}$ is between 0.167 $MW_{th}$ and 0.231$MW_{th}$, which corresponds to an error on the overall demand between 2.39 % and 3.22 %.

The model makes it possible to predict a trend of thermal losses to ground (Fig. 5) similar to that shown in Fig. 2, even if the percentage values reported in Fig. 5 are on average lower than those reported in Fig. 2. However, the error in the forecast of the incidence of losses cannot be assessed with certainty, as values reported in Fig. 2 are not obtained from direct measurements [32].

**Table 3**
Numerical values of grey box DH model parameters with admitted range of variation.

| Parameter | Admitted variation range | Training result |
|---|---|---|
| $C_{DH}$(MJ K$^{-1}$) | from 1257 to 3352 | 3352 |
| $K_{SH}$ (kW K$^{-1}$) | from 100 to 400 | 282 |
| $\dot{Q}_B$(kW) | from 1000 to 1500 | 1000 |
| $K_L$(kW K$^{-1}$) | from 8 to 8.5 | 8 |
| $f_c$(kW K$^{-1}$) | from 0 to1500 | 1500 |

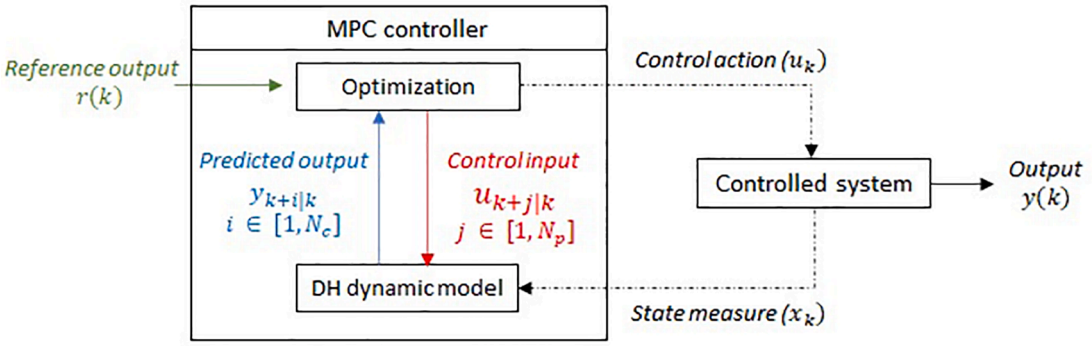


**Fig. 4.** Scheme of the MPC control.

**Table 4**
RMSE values (grey box DH model): training and testing data set.

| Data set | $T_r$(°C) | $\dot{Q}$($MW_{th}$) |
|---|---|---|
| Training data set (first 3 weeks of January 2018) | 1.59 | 0.219 |
| Testing (22 January/4 February 2018 and 22 March/4 April 2018) | 1.75 | 0.231 |
| Testing all data (year 2018) | 1.54 | 0.167 |

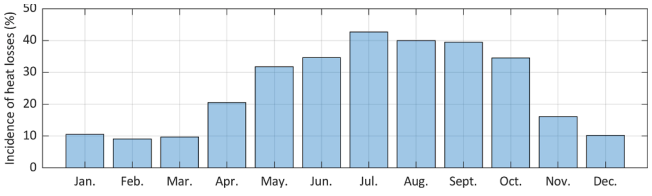


**Fig. 5.** Monthly percentage impact of thermal losses on demand (grey box model).

As for the settings of the optimization problem, these are described in the following points:

- The return temperature is kept close to the value of 63 °C ($r(k)$ which also coincides with $T_{r,sp}$ in Equation (1)) with an allowed range variation between 59° C and 68 °C.
- Total demand $\dot{Q}$ is inferiorly bound to the satisfaction of user demand. This constraint can be expressed by:

$$\dot{Q} \geq \dot{Q}_U \tag{10}$$

- The water flow rate $\dot{m}_{DH}$ is maintained between 5 kg $s^{-1}$ and 70 kg $s^{-1}$ compatibly with the specifications of the installed circulating pumps.
- The supply temperature $T_s$ is bound to different intervals in relation to the season of the year. For the winter months (November to March) it can vary from 85 °C to 95 °C if the outside temperature is greater than 1 °C. Otherwise the allowed range is 90–95 °C. For the mid-season months (April–May and September–October), $T_s$ can vary between 78 °C and 95 °C, while for the summer months (June-August) the permitted range is 75–95 °C. These values have been chosen compatibly with the operating limits of the plant, as already described in subsection 2.1.3.
- From the observation of time variation of measured data, a maximum rate of change has been set for the decision variables: 20 kg $s^{-1}$ for the water flow rate and 5 °C for the supply temperature.
- The prediction horizon is set to 1 h (4-steps) and the control horizon is set to 1 step. Although it is a high thermal inertia system, the choice of a 1-hour prediction horizon was made with a view to the accuracy of the MPC. Indeed, with a long prediction horizon, the reliability of the forecast can be low, and this compromises the control planning ability. Whereas, by updating the initial conditions of the MPC at every hour with real system measurements, the probability of having important deviations between the model prediction and the real plant behaviour decreases.

It is important to point out that the model used to formulate the MPC is based only on energy balance (thermal model). It does not model, at present, the pressure losses in the pipes of the DH network. In practical application, the flowrate adjustment should take this aspect in consideration in order to ensure the required pressure levels.

### 2.3. Reinforcement Learning

Reinforcement Learning (RL) has a long history in the artificial intelligence research field [20], but only in recent years, with the adoption of deep neural networks, this framework has been adopted extensively in real applications. The common objective in most RL problems is to find a policy which maximizes the reward w.r.t a specific goal.

In this work, a standard RL setting where the agent interacts with the environment over a number of discrete time steps is considered. The environment can be seen as a Partially Observable Markov Decision Process (POMDP) in which the main task of the agent is to find a policy π that maximizes the expected sum of future discounted rewards:

$$G_t = \sum_{k=t+1}^{T} \gamma^{k-t-1} Rw_k \tag{11}$$

where $\gamma \in [0, 1)$ is the discount factor and $Rw_k$ is the reward at time $k$, given the state $s_t$ and the action$a_t \sim \pi(|S_t)$. The Markov Decision Process (MDP) is partially observable, in the case of visual navigation, because in every step the agent has no access to the true state $S_t$ of the environment, but only to an observation $o_t$ of it.

For all the tasks, the actor-critic framework is used. It divides the agent into two components: (i) actor is the one that interacts with the environment and learns to perform actions. It collects a series of trajectories, composed by observations, actions, and rewards. (ii) Critic has the role of evaluating the actor performances. It uses these trajectories to learn to estimate the expected sum of the future discounted rewards of the actor policy $G_t$.

The critic can estimate $G_t$ by computing one of the following two functions: the value function:

$$v_\pi(s) = \mathbb{E}_\pi[G_t|S_t = s] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k Rw_{k+t+1} \middle| S_t = s\right], for\ \ all\ \ s \in S \tag{12}$$

or the action-value function:

$$q_\pi(s,a) = \mathbb{E}_\pi[G_t|S_t = s, A_t = a] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{k+t+1} \middle| S_t = s, A_t = a\right] \tag{13}$$

which describes the expected return after choosing an action $a_t$ in state $S_t$ and thereafter following policy $\pi$.

#### 2.3.1. Parameters and settings

As for the MPC, also the RL is formulated to replace low-level control (subsection 2.1.3): the control actions are therefore the water supply temperature and the value of the flow in circulation. As for the MPC, the pressure levels of the circuit are not taken into account and the flow rate regulation is performed by the control only to accomplish the energy balances.

The observation $o_t$ is composed of 4 signals, the hour of the day, the external air temperature $T_e$, the difference between the return temperature estimated by Equation (1) with respect to the set point temperature ($T_r - T_{r,sp}$) and the difference between the heat demand estimated by Equation (4) and that measured at previous step. The measured heat demand at the actual time step can be considered as a good feature to predict the heat demand at the next step (15 min) since the best fit rate of the difference $y_{k+i|k} - y_{k|k}$ is 89 % and the RMSE is equal to 0.226 $MW_{th}$ in the year 2018. As reward function is considered the minimization of the overall thermal demand and an Actor-Critic (AC) agent is employed for training. The AC agent uses a model-free, online, on-policy reinforcement learning method, to implement actor-critic algorithms, such as A2C and A3C. The observation space is considered continuous, and the action space is considered discrete where the supply temperature action space $\in \pm[0, 0.5, 1, 2, 3, 4, 5]$ and the flow rate action space $\in \pm[0, 0.5, 1, 1.5, 2, 3, 4, \cdots, 20]$ for a total of 585 combinations. For the critic and actor networks is considered the same deep neural network architecture composed of 8 layers: input layer, fully connected layer,

hyperbolic tangent activation layer, fully connected layer, hyperbolic tangent activation layer, fully connected layer, rectified linear unit layer, output fully connected layer. The size of the fully connected layer and hyperbolic tangent is set at 32.

## 3. Results and discussion

In this Section the results of the application of MPC and RL to the CHP-DH plant are presented. The first two subsections (3.1 and 3.2) describe the expected operation of the system when controlled respectively with the MPC and the RL. The observed dynamic of the plant with the two controls are compared with the current operation of the plant (i.e., real data), which will be considered as the reference case. In both cases the application of the control is evaluated in a simulated environment and the controlled system is represented by the grey box DH model presented in subsection 2.2.1. However, both controls are applicable to the actual CHP-DH plant as they are characterized by very short execution times (i.e., milliseconds). To generalize the analysis, subsection 3.3 contains a critical discussion of the advantages and disadvantages that emerged in the modelling of the two controls for the system under study.

### 3.1. Application of Model Predictive Control

The results of the MPC are divided in two parts. The first (subsection 3.1.1) describes the operational evaluation and the performance obtained with reference to the current operation of the CHP-DH plant. In the second part (3.1.2), the limitations and observations that emerged during control modelling are critically discussed.

#### 3.1.1. Assessment of the control in operation

In order to show the results of the application of the MPC to the CHP-DH plant, two reference periods were considered: one representative of the winter season (from 7th to 21st of January) and one of the mid-season (from 18th of March to the 1st of April). A summer period has not been evaluated as the control strategy that would evaluate the MPC would be similar to the current one, in which the supply temperature is already kept close to the minimum limit (75 °C).

Regarding the winter period, Fig. 6 compares the total thermal demand covered by the heat generator in presence of the MPC with respect to the reference case. It can be noted the potential effectiveness of the MPC in reducing the thermal demand, especially in the hours of low heat demand. To make the distinction between day and night periods clearer, in Fig. 6 the central part of the day (from 7.00 am to 8.00 pm) is highlighted in light grey. The distinction between the time bands in Fig. 6, as for the other figures in Section 3 (Figs. 7, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 21, 22, 23, 24), is not actually considered in the optimal control, but is used here only to facilitate the representation of the results and make the comparison with the current control more direct.

A lowering of the minimum load is observed with the MPC. This advantage is most evident when considering the night hours (from 8.00 pm to 7.00 am, Fig. 6). The base load, in fact, goes from 1.6 $MW_{th}$ in the reference, to 1.36 $MW_{th}$ in the case of MPC, with a reduction of about 15 %. This lowering is caused by the different regulation strategy that the

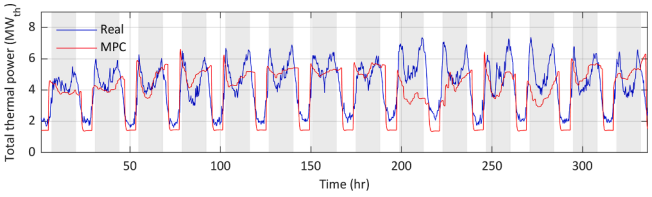


**Fig. 6.** Total thermal demand covered by heat generator (CHP engine and boilers) in a winter reference period (7th-21st of January) comparison between real data and MPC.

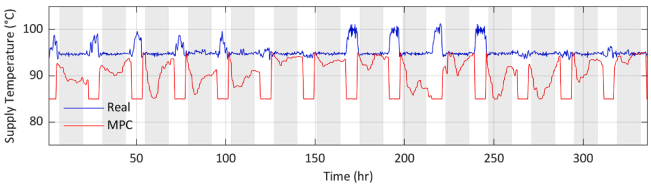


**Fig. 7.** Water supply temperature in a reference winter period (7th-21st of January): comparison between real data and MPC.

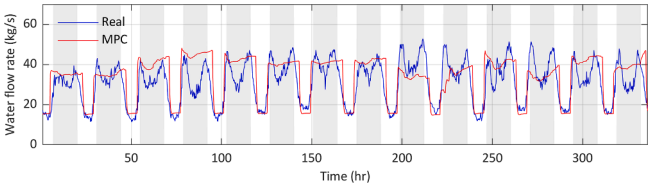


**Fig. 8.** Water flow rate in a reference winter period (7th-21st of January): comparison between real data and MPC.

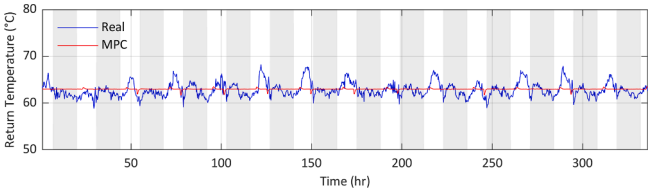


**Fig. 9.** Water return temperature in a reference winter period (7th-21st of January): comparison between real data and MPC.

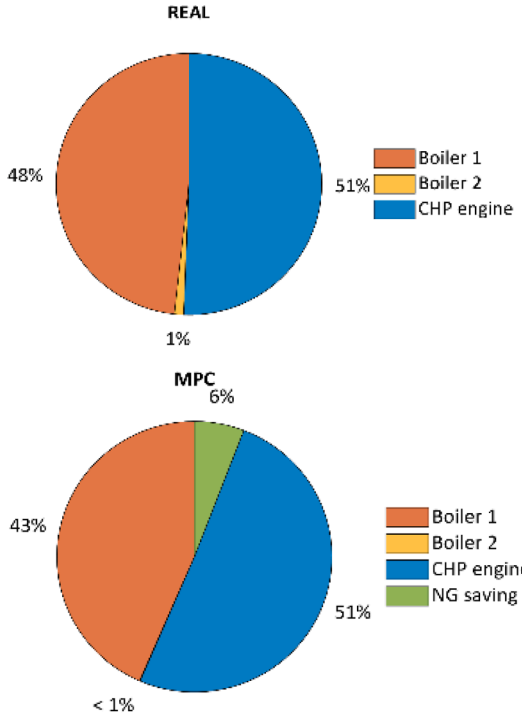


**Fig. 10.** Percentages of NG of boilers and CHP engine (7th-21st of January): comparison between real data and MPC results (percentages refer to real data).

MPC sets in terms of supply temperature and water flow rate. After 8.00 pm the supply temperature is kept for almost all the time at the minimum value allowed by the constraint (85 °C, as described subsection

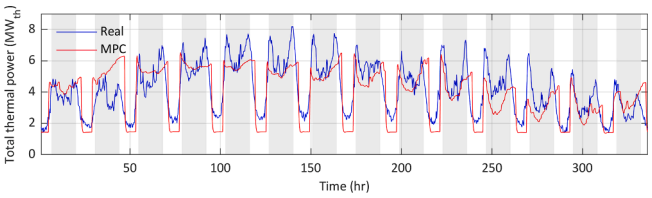


**Fig. 11.** Total thermal demand covered by heat generator (CHP engine and boilers) in a mid-season reference period (18th March–1st April): comparison between real data and MPC.

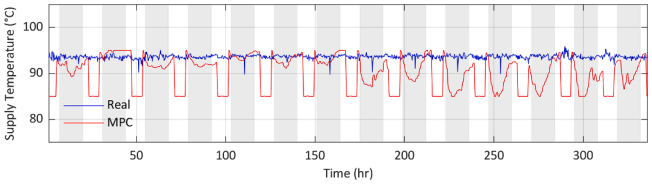


**Fig. 12.** Water supply temperature in a mid-season reference period (18th March–1st April): comparison between real data and MPC.

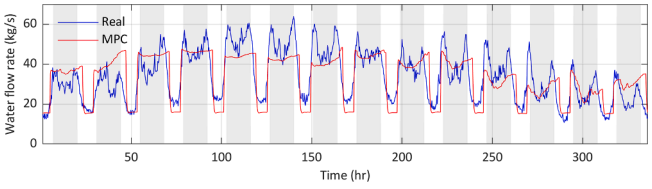


**Fig. 13.** Water flow rate in a mid-season reference period (18th March–1st April): comparison between real data and MPC.

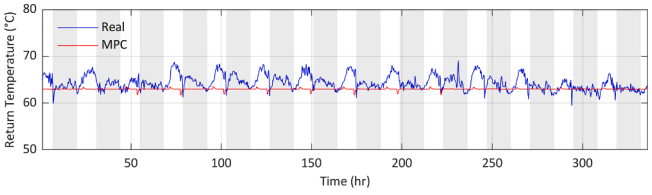


**Fig. 14.** Water return temperature in a mid-season reference period (18th March–1st April): comparison between real data and MPC.

2.2.3), and the water flow rate is more or less similar to the reference case (areas not highlighted in the Figs. 7 and 8). This is possible because, at night, thermal losses, which depend heavily on water temperatures, represent an important contribution in the total heat demand: they are almost the 18 % of night heat demand (the share drops to 11 % during the day). The reduction of the estimated losses towards the ground at night is about 5.2 %, going from 93.8 $MWh_{th}$ to 88.9 $MWh_{th}$. However, MPC effectively lowers the impact of losses even during the day, which pass from 94.1 $MWh_{th}$ to 91.6 $MWh_{th}$. Also in this case, albeit less evidently than at night, the MPC tends to lower the supply temperature, leaving the flow rate values close to those of the reference case (areas highlighted in light grey in Figs. 7 and 8). Indeed, during the day, the MPC tries to lower the supply temperature as much as possible, following the dynamics of the users' heat demand.

As described in Section 2.2, the MPC has also been formulated to keep the water return temperature close to an imposed value ($T_{r,sp}$ in Equation (1)). Fig. 9 compares the trend of the water return temperature obtained with the MPC with the reference case (i.e., measured data). As can be seen, the MPC is able to maintain the required temperature level (63 °C) more punctually than the actual control.

To summarize, Table 5 contains an estimate of the main energy

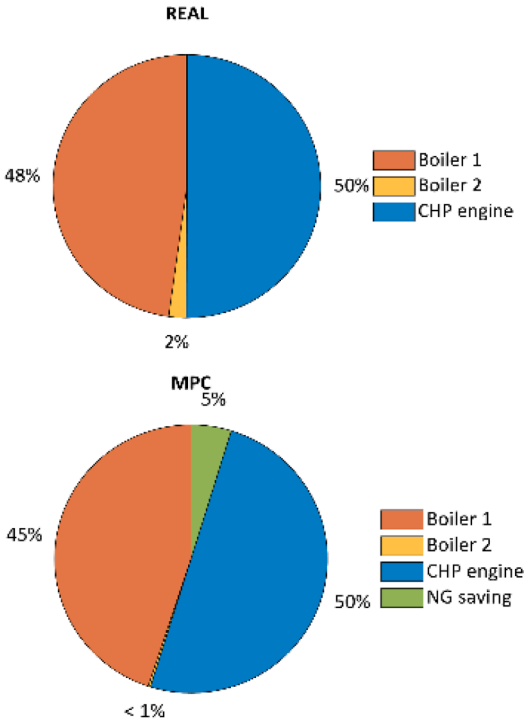


**Fig. 15.** Percentages of NG of boilers and CHP engine (18th March – 1st April): comparison between real data and MPC results (percentages refer to real data).

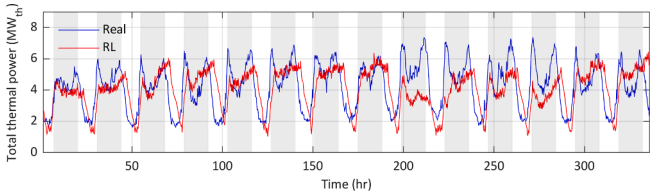


**Fig. 16.** Total thermal demand covered by heat generator (CHP engine and boilers) in a reference winter period (7th-21st of January): comparison between real data and RL.

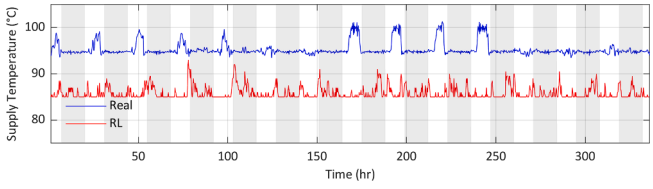


**Fig. 17.** Water supply temperature in a reference winter period (7th-21st of January): comparison between real data and RL.

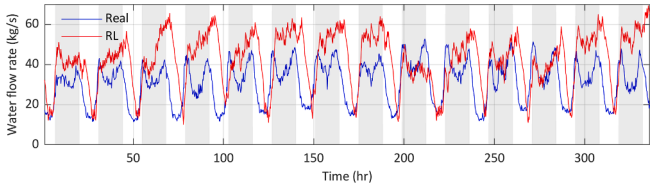


**Fig. 18.** Water flow rate in a reference winter period (7th-21st of January): comparison between real data and RL.

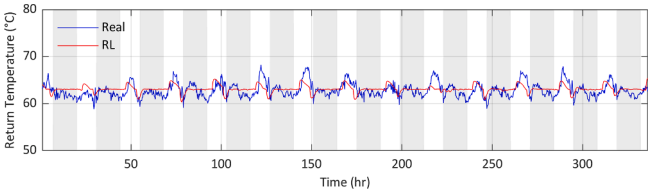


**Fig. 19.** Water return temperature in a reference winter period (7th-21st of January): comparison between real data and RL.

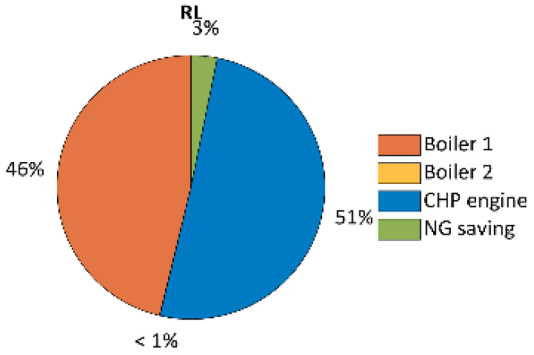


**Fig. 20.** Percentages of NG of boilers and CHP engine (7th-21st of January) with RL (percentages refer to real data).

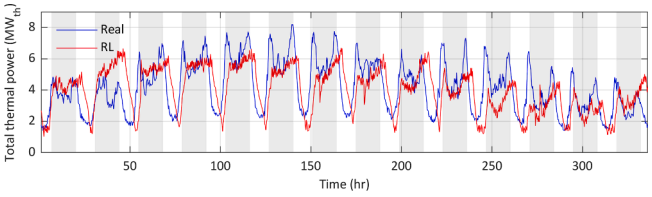


**Fig. 21.** Total thermal demand covered by heat generator (CHP engine and boilers) in a reference mid-season period (18th March – 1st April): comparison between real data and RL.

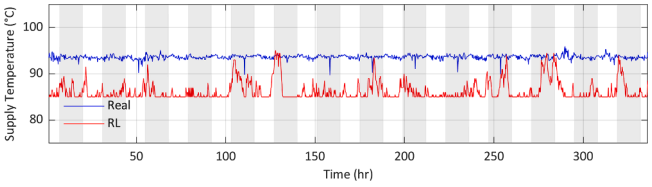


**Fig. 22.** Water supply temperature in a reference mid-season period (18th March – 1st April): comparison between real data and RL.

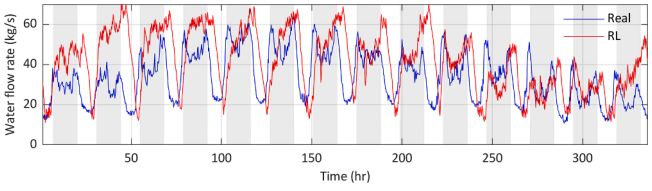


**Fig. 23.** Water flow rate in a reference mid-season period (18th March – 1st April): comparison between real data and RL.

parameters in the period analyzed. As can be seen, the MPC allows to obtain a reduction of 8.2 % of the total heat demand that the CHP engine and the boilers must cover, of which about 7 % derives from the reduction of the thermal losses towards the ground (7.3 MWh$_{th}$). In

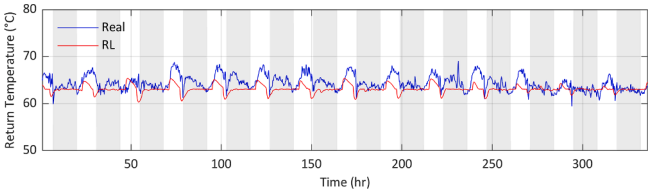


**Fig. 24.** Water return temperature in a reference mid-season period (18th March – 1st April): comparison between real data and RL.

**Table 5**
Energy estimates and NG consumption (7th–21st of January): comparison between real data and MPC.

| Data | Real | MPC |
|---|---|---|
| Total thermal demand (MWh$_{th}$) | 1413 | 1297 |
| Thermal losses (MWh$_{th}$) | 187.9 | 180.6 |
| NG consumption (Sm$^3$) | 218,580 | 205,780 |

general, the thermal losses are reduced by 3.9 % compared to the current control: reduction of 2.6 % considering the daily hours and 5.2 % considering the night hours. Considering the efficiency of the CHP engine and boilers shown in Table 1 and a calorific value for NG equal to 9.4 kWh Sm$^{-3}$, Table 5 also reports the estimation of the fuel savings that can be obtained with the MPC. In the period analysed (7th-21st of January), it is obtained a potential NG saving of 12,800 Sm$^3$, corresponding to a primary energy saving of about 126 MWh and to avoided emission of 23.3 tCO$_2$ (values calculated considering a conversion factor in primary energy of natural gas of 1.05 [36] and an equivalent emission factor of 0.1936 kgCO$_2$/kWh [37]). Going into detail, Fig. 10 shows the NG use share of the single technologies involved in satisfying the thermal demand. The percentages given in the pie chart are referred to the total consumption for the reference case. In this way, there is an additional slice in the case of MPC which quantifies the savings. The reduction in NG consumption is due exclusively to the reduction in boilers consumption. This is a favorable aspect for the operation of the heat generation plant as it leads to an increase in the share of demand satisfied by the CHP engine.

It is interesting to note that, the evaluation of the results shown in Table 5 and in Fig. 10 is made only by knowledge of the performance curves of the generator. This makes the analysis generalizable also to different generation systems as long as the performance curves are known.

The analysis on the winter period confirms the effectiveness of the MPC on the CHP-DH management. These observed performances were also confirmed by the analysis performed on the mid-season reference period (from 18th of March to the 1st of April). Fig. 11 represents the comparison between the total thermal demand in the case of MPC and the reference case in the chosen period. Also, in this case there is a reduction in the basic demand: which goes from 1.40 MW$_{th}$ to 1.37 MW$_{th}$. It is confirmed that the tendency of the MPC is to decrease the water supply temperature (Fig. 12) as much as possible at night, while the flow rate tends to follow the measured trend (Fig. 13). However, it is observed the trend of the MPC to mitigate the variation in demand which fluctuates in a smaller interval compared to the real data. Furthermore, observing the return temperature (Fig. 14), the considerations made for the winter period can be repeated.

Also in the mid-season, the energy assessment allows to evaluate the effectiveness of the MPC. The total demand covered by the generator decreased by 6.64 %, going from 1385 MWh$_{th}$ to 1293 MWh$_{th}$, while the estimated thermal losses towards the ground decreased by about 3.5 %, going from 188.6 MWh$_{th}$ to 182 MWh$_{th}$ (Table 6). Again, the highest percentage reduction of the estimated thermal losses occurs at night. A reduction of about 4.6 % at night was estimated, the percentage become 2.8 % in the hours of the day. Even in the case of mid-season, the MPC

**Table 6**
Energy estimates and NG consumption (18th March–1st April): comparison between real data and MPC.

| Data | Real | MPC |
|---|---|---|
| Total thermal demand ($MWh_{th}$) | 1385 | 1293 |
| Thermal losses ($MWh_{th}$) | 188.6 | 182 |
| NG consumption ($Sm^3$) | 213,310 | 203,130 |

allows to achieve a saving of 10,180 $Sm^3$ of NG (Table 6), corresponding to a primary energy saving of about 100.5 MWh [36] and to avoid the emission of 18.5 $tCO_2$ [37]. Moreover, observing the composition of the technologies involved in covering the demand (Fig. 15), the observed trends of fuel saving in Fig. 10 are confirmed. As for the winter case the saving of NG happens mainly at the expense of the two boilers that reduce their hours of operation. This leads once again to an increase in the percentage of demand met by the CHP.

#### 3.1.2. Critical analysis of the control technique

The results of applying the MPC to the CHP-DH plant showed the potential effectiveness of the control in optimizing the energy consumption of the plant. However, it is important to highlight that the benefits obtained in quantitative terms must be considered as potential and not actual performance. Observing the results in terms of the objective pursued, it is evident the ability of the MPC to evaluate the optimal solution in compliance with the constraints. However, it is not possible to simply exclude inaccuracies in the prediction capacity of the model during the operation. Such inaccuracies may be due either to an inaccurate formulation and/or configuration of the model or to the presence of not expected disturbances. Both aspects also clearly emerged during the MPC modelling. Indeed, the identification of the model configuration for the case study involved a great effort. As can be seen in the formulation described in Section 2.2.1 (Equation (1)), it was necessary to add a corrective factor to represent the limitation in the variation of the return temperature.

The observation of the comparison of the return temperatures obtained with the MPC with the measured data (Figs. 9 and 14) also allows to note that the model is not able to capture the instant dynamics of the system. The trend of the measured return temperature in fact contains high frequency oscillations of which it is difficult to predict the cause. These are probably due to unexpected causes (e.g., delays in control actions, measurement errors or mechanical systems with delay) that it was not possible to frame with a mathematical representation. These disturbing contributions are not successfully represented by the model.

There is also one last aspect that is worth highlighting: the assessment of the thermal losses to the ground should be seen in qualitative rather than quantitative terms. Indeed, the accuracy of the quantification of thermal losses towards the ground is not certain. Since the reference is not derived from direct measurements, it is not possible to precisely quantify the reliability of the model in this sense. This should be considered when modelling controls whose formulation requires data. In fact, detailed measurements of each physical size of interest are not always available.

### 3.2. Application of Reinforcement Learning

Also for RL the presentation of the results is divided into two parts. In the first, there is the description of the operating behaviour of the CHP-DH plant in comparison to the reference case (3.2.1). In the second (3.2.2) the discussion on the limitations that emerged during RL modelling is provided.

#### 3.2.1. Assessment of the control in operation

As in the case of the MPC, the results for the RL are also evaluated in the two representative periods in winter (from 7th to 21st of January) and mid-season (from 18th of March to the 1st of April). For the reasons given at the beginning of the previous subsection, also for the RL the analysis of the results will exclude the summer case.

Fig. 16 shows the comparison between the total heat demand obtained with the RL and the reference case (also in this case the hours of the day are highlighted) in the winter period. It can be observed that the RL can effectively meet the thermal demand of the users minimising the thermal losses. With the RL there is a considerable reduction (about 35 %) of the base load, which becomes equal to 1.035 $MW_{th}$. Figs. 17 and 18 show respectively the supply temperature and flow rate trends established by the RL in comparison to the reference case. It can be noticed that, compared to the MPC (Figs. 7 and 8), the RL tends to suggest control strategies that include higher flow rates and as low as possible supply temperatures, always compatible with the operational constraint. In the MPC the dynamics of the users' heat demand were more followed by the supply temperature trend (Fig. 7), while in the RL it is the flow rate that reflects the dynamic changes in the demand (Fig. 18). In addition, the RL is more capable than the MPC in avoiding the power peak (Fig. 16) occurring between the shift from the night to the day.

By suggesting lower supply temperatures, the RL reduces losses by about 6.54 % over the period considered. At night only, the reduction in losses is about 7 % from 93.8 $MWh_{th}$ to 87.2 $MWh_{th}$. During the day it is about 6 % (from 94.1 $MWh_{th}$ to 88.4 $MWh_{th}$). Fig. 19 shows the trend of the water return temperature. It can be observed that also with the RL the control keeps the return in a range rather close to the imposed set point. However, there is a trend in the temperature variation more in accordance with the measured data.

Table 7 shows the energy and fuel consumption assessed for the RL in comparison with the reference control. While losses appear to be decreasing more than with the MPC, the RL produces a lower reduction in demand: around 4.4 %. The overall NG consumption is therefore reduced less than the in the case with the MPC. Fig. 20 shows the NG demand share of the technologies involved in satisfying demand. This was evaluated with the efficiencies of the technologies reported in Table 1 and considering a calorific value of 9.4 kWh $Sm^{-3}$ for the NG. An overall saving of 3 % (6910 $Sm^3$ of NG) is observed, while with the MPC it reached 6 % in the same period. However, although the saving seems to be low in percentage, the RL allows to save about 68 MWh of primary energy [36] and to avoid the emission of 12.6 $tCO_2$ in the period analysed [37]. Furthermore, also in this case, the saving is made only at the expense of the boilers, confirming the advantage of control in increasing the percentage share of the demand covered by the CHP engine.

Considering the mid-season period (from 18th of March to the 1st of April), the control is less effective than in the winter case is observed. Fig. 21 shows the total heat demand in the period considered in relation to the measured data. The RL produces a reduced decrease of the base load from 1.4 $MW_{th}$ to 1.02 $MW_{th}$. Although less evident than in winter, the tendency of the RL is to adjust the demand more with the flow rate (Fig. 23) than with the supply temperature (Fig. 22). In this case, however, the control is not able to evaluate solutions with supply temperature always close to the lower limit. However, the RL confirms in the observed period its ability to reduce thermal losses. The RL allows to avoid 6.2 $MWh_{th}$ of thermal losses in the night hours and about 5.4 $MWh_{th}$ in the hours of the day. Moreover, as shown in Fig. 24, the same considerations for the return temperature apply as for the winter case.

Table 8 shows the values of the total heat demand covered by the heat generator, the total thermal losses towards the ground and the total

**Table 7**
Energy estimates and NG consumption (7th–21st of January): comparison between real data and RL.

| Data | Real | RL |
|---|---|---|
| Total thermal demand ($MWh_{th}$) | 1413 | 1351 |
| Thermal losses ($MWh_{th}$) | 187.9 | 175.6 |
| NG consumption ($Sm^3$) | 218,580 | 211,670 |

**Table 8**
Energy estimates and NG consumption (18th March–1st April): comparison between real data and RL.

| Data | Real | RL |
|---|---|---|
| Total thermal demand ($MWh_{th}$) | 1385 | 1338 |
| Thermal losses ($MWh_{th}$) | 188.6 | 177 |
| NG consumption ($Sm^3$) | 213,310 | 208,140 |

consumption of NG in comparison with the measured data. In this case the saving is about 5170 $Sm^3$ of NG which correspond to about 51 MWh of primary energy [36] and 9.4 $tCO_2$ avoided [37]. Fig. 25 instead shows the consumption of NG for the technologies involved. Fig. 25 is referred to the case in which the plant is controlled with the RL and, to highlight the NG saving, the percentage are referred to the reference case (Fig. 15). Although with a lower impact, this case confirms the fuel savings at the expense of boilers.

*3.2.2. Critical analysis of the control technique*

The evaluation of the results with the application of the RL to the CHP-DH plant also showed the potential of the control to reduce the impact of the thermal losses. However, as for the MPC, observations should be made on the assessment of the reliability of the results. As already mentioned in the methodology (Section 2.3), the RL used in this work was formulated with a double target to be pursued: the return temperature and the net demand of users (excluding losses). The latter, in the absence of measured data, was evaluated with the model of the controlled system. This is a first critical point of modelled RL, which highlights the difficulty of formulating a purely data-based control when some measures are not available. Implementing a purely RL control requires in fact a large dataset for training. This dataset should contain both breakdowns of measures (as in our case the net demand compared to the total covered by the heat generator) and all the possible configurations in which the plant can be found to operate. Furthermore, it would also be very useful to shave operating points where the plant violates operational constraint. In RL constraints are usually defined as penalties and then included in the reward; it is also a tricky matter to adjust the weights of penalties. Therefore, the process of identifying the reward function requires information about the right behaviour of the plant to be optimized (e.g., tracking of a setpoint, minimization of energy cost) but also what the plants should not pursue (e.g., violation of constraints). These data are not only often not available but would involve having to operate the plant under intentionally unfavourable conditions for a time that is difficult to estimate.

The difficulty of planning the duration of the time periods required to obtain sufficient measurements and/or to set the best function is another aspect to consider in the case of RL. In our case, for instance, the adjustment was performed using a trial-and-error approach. The training process was quite slow and complex. Several hours and several trials were needed to set the optimal weights of the reward function,

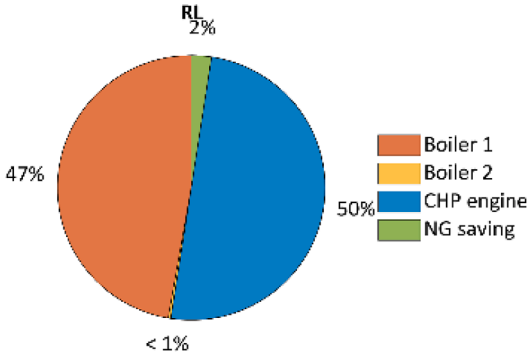


**Fig. 25.** Percentages of NG use of boilers and CHP engine (18th March – 1st April) with RL (percentages refer to real data).

with the conclusion that it is not possible to exclude that there are better configurations than the one presented in this paper. Consequently, the control logic presented in the previous Section (3.2.1), in which it is noted that the strategy chosen by the RL tend to prefer the management of the flow rate rather than the supply temperature, is strongly affected by the RL training solution. It is therefore not obvious to find a physical interpretation of this choice preferred by the control.

### 3.3. General comparison of the pros and cons of MPC and RL

The previous Sections show the results obtained from the application of the two controls to the CHP-DH plant under study. Both controls have been found to be potentially effective in reducing the energy impact of the plant compared to its current operation. However, as already discussed, some important differences between MPC and RL have emerged during their formulation and application. In this concluding Section, based on the results described in Sections 3.1 and 3.2, the key differences between the two controls will be listed. In particular, the critical points and advantages will be discussed with the aim of increasing the awareness on pros and cons of advanced control techniques. The discussion will be structured by addressing three fundamental aspects: (i) main drawback of the controls, (ii) difficulty of mathematical formulation and (iii) ability to adapt to system changes.

(i) Main drawback of MPC and RL.

In general, both the MPC and the RL have a critical point that represents the main barrier to overcome to ensure the desired performance. For the MPC the critical point is the system model, while for the RL it is the training process.

The effectiveness and reliability of the MPC are closely related to the accuracy of the forecast model. Two levels of difficulty need to be addressed: (i) identifying the right model architecture without exceeding complexity and (ii) having enough system-related information and/or measured data. Furthermore, the model should not only be accurate in replicating the known dynamics of the system but should also be able to predict possible responses of the system when subjected to stresses or inputs other than those known. For complex systems that are affected by many factors, even difficult to predict, such as in the CHP-DH plant considered, this latter aspect is not always taken for granted. In our case, in fact, the main difficulty that was encountered in the formulation of the MPC was precisely in defining the model of the DH network. Even because the data measured are rarely available in the required completeness and detail.

On the other hand, the RL, which generally does not need system knowledge, requires considerable effort to get through the training process. For complex systems, such as that seen in this work, the training process of the RL is very expensive and could almost certainly lead to phases, even long, in which the controlling actions generate high system malfunctions (trial and error approach). This freedom to make mistakes that must be granted to the control is not always acceptable for energy applications as for the case study analysed. For this reason, as in the formulation presented in this paper, it seems practically obligatory to perform «offline» training a priori and make online only the refinement of the solution (e.g., to adapt it to the weather conditions and the loads of the network that can change). The latter aspects reduce the independency of the control from having information on the system.

(ii) Difficulty of mathematical formulation.

Beyond the drawback represented by the model (i), the differences in terms of mathematical formulation between MPC and RL are considerable. MPC in fact is one of the most promising and most used control techniques, even in the industrial field, due to its effectiveness and ease of formulation. The main advantages of the MPC from this point of view are:

- It always allows to evaluate the optimal solutions (this is clear from the results for the CHP-DH plant, Tables 5 vs 7 and Table 6 vs 8).

- Its mathematical formulation is quite simple. In fact, given the reliability of the model for granted, it is possible to define rather clear control target.
- It allows to easily consider physical constraints to be imposed on the control, both soft and hard.

On the other hand, the aspects that represent advantages for the MPC are weaknesses for the RL. In fact, the RL:

- Does not allow to evaluate the absolute optimal solution.
- Its mathematical formulation is complex. In general, RL requires good computational skills. Even with a trained offline control, as in the case presented, many computational difficulties and very long times of setting to identify the hyperparameters, to define the function of reward and the deep neural network architecture for the actor-critic representation (i.e., how many neurons, how many layers, what neuron typology, etc.) are required.
- With the RL there are no guarantee that hard constraints are satisfied. This aspect did not clearly emerge from the results we obtained, as the identified reward proved to be particularly effective in effectively controlling the CHP-DH plant. However, it should be borne taken in mind that in RL constraints are usually defined as penalties and then included in the reward: in this way, the agent learns to avoid such penalties. This does not guarantee, as in the case of the MPC, the guarantee of compliance with the constraints, but correlates everything with the definition of the reward function.

(iii) Ability to adapt to system changes.

The aspect of the adaptability of the MPC and the RL to changes in configuration or operating modes of the system has different characteristics if we consider this ability as a feature to have online or offline. Indeed, it is quite simple to make changes to the MPC structure. Therefore, it allows to easily include modifications and new features offline. Changes in the structure of the RL are to be avoided as they would involve a new training process with all the difficulties attached and its onerousness. In general, however, a reformulation of the structure of the RL is possible but requires the whole offline training. On the other hand, the MPC has a fixed structure once implemented. The RL does not have a fixed structure and has the advantage over the MPC to be able to improve during its operation. Therefore, from the continuous collection of measured data the RL has the ability to evolve considering unexpected changes in the actual operation (ability to adapt online).

In conclusion, Fig. 26 summarizes the main findings of the performed comparison. The objective of the paper is not to suggest one type of control rather than the other, but to provide a case study that can be taken as a reference for anyone who wants to implement one of these techniques to control an energy system.

## 4. Conclusions

This work showed the application of two advanced control techniques, namely Model Predictive Control (MPC) and Reinforcement Learning (RL) to a case study consisting of a Combined Heat and Power plant serving a District Heating network (CHP-DH) located in central Italy.

The work has a twofold objective. Firstly, to evaluate the effectiveness and potential of these control techniques in the CHP-DH plant considered. This should be understood as a preliminary investigation to evaluate the replacement of the current control technique with a more advanced one. The second objective, which derives from the realization of the first, consists in presenting a critical analysis of the pros and cons encountered during the modelling of the two controls for the CHP-DH system on a real case study, with the intention of being able to offer food for thought for anyone else who wants to apply these two control techniques to large-scale operating systems.

Based on the knowledge of the CHP-DH plant characteristics and the availability of data measured over a whole year, the two controls (MPC and RL) were formulated and tested in a simulated environment. In both cases, the purpose of the control is to be able to satisfy the demand of the users, reducing the energy expenditure for network losses. To do this, both the supply temperature and the circulating flow rate have been used as control variables.

The main results and considerations that emerged can be summarized in the following points:

| | MODEL PREDICTIVE CONTROL | REINFORCEMENT LEARNING |
|---|---|---|
| (i) Main criticality | Its effectiveness depends heavily on **goodness of the model** (difficulty in predicting unexpected disturbances). | Onerous **training process.** It is necessary that the plant violates any constraints. |
| (ii) Matematical formulation | • Easier to formulate.<br>• Optimal solution guarantee.<br>• It handles explicitly hard constraints in the optimization problem. | • Complex formulation (difficulty in formulating the individual parts of the control).<br>• No guarantee of optimal solution.<br>• No guarantee that hard constraints are satisfied. |
| (iii) Adaptability | • It does not have a fixed structure (modifications and new features are easily included).<br>• There is no chance that it will improve with operation (online). | • It could improve itself through data (online).<br>• It does not have a fixed structure. |

**Fig. 26.** Key aspects emerged from the comparison of MPC and RL.

- Both controls showed their effectiveness in meeting the heat demand of the users, reducing the heat losses of the network. In a reference winter period, thermal loss reductions of up to 3.9 % were achieved for the MPC and up to 6.54 % with the RL. During the reference period (2 weeks) with the MPC, avoidance of $CO_2$ emissions up to 23.3 $tCO_2$ and up to 12.6 $tCO_2$ with the RL were assessed.
- Although the same objective was achieved, the two controls showed different trends in the selection of control actions. The MPC in fact showed a reduction in the supply temperature on average less evident in the hours of greatest demand. The RL, on the other hand, showed more of a propensity to lower the supply temperature and to act on the flow rate to chase up increases in demand from users.
- Both in the MPC modelling and in the RL, critical issues emerged. In the first case, the reliability of the results is very linked to the model used in the control. If there are not many data available and there are disturbances not easily identifiable, the MPC can suggest strategies whose impact in reality is different from the estimated one. On the other hand, the RL, which is not model based, is potentially more adaptable to unexpected variations. However, RL requires a number of computational skills for its formulation that cannot be overlooked.

These are the main considerations that emerged from the application of the controls to the case study. What has emerged is that the MPC is certainly the control technique to be preferred when the physics of the problem is known, the optimal solution is required and the compliance with hard constraints is essential. On the other hand, the RL can be applied, albeit with a number of difficulties, even when it is not possible to have any knowledge of the system. Moreover, this is the control to be preferred when it is required a controller able to improve itself through data gathering during operation.

## CRediT authorship contribution statement

**A. Mugnini:** Conceptualization, Methodology, Software, Validation, Writing – original draft, Formal analysis. **F. Ferracuti:** Conceptualization, Methodology, Software, Validation, Writing – review & editing. **M. Lorenzetti:** Visualization, Supervision. **G. Comodi:** Methodology, Supervision, Project administration, Funding acquisition. **A. Arteconi:** Conceptualization, Methodology, Writing – review & editing, Visualization, Supervision.

## Declaration of Competing Interest

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

## Acknowledgements

This study has received funding from European Union's Horizon 2020 Research and Innovation programme under grant agreement No 824441 (MUSE GRIDS).

## References

[1] European Commission. 2050 long-term strategy. Available online: https://ec.europa.eu/info/energy-climate-change-environment/implementation-eu-countries/energy-and-climate-governance-and-reporting/national-long-term-strategies_en (accessed on 30 April 2022).
[2] International Energy Agency (IEA). Heating. Available online: https://www.iea.org/reports/heating (accessed on 30 April 2022).
[3] Connolly D, Lund H, Mathiesen BV, Werner S, Möller B, Persson U, et al. Heat Roadmap Europe: Combining district heating with heat savings to decarbonise the EU energy system. Energy Policy 2014;65:475–89. https://doi.org/10.1016/j.enpol.2013.10.035.
[4] Department of Development and Planning AU. EnergyPLAN. Heat Roadmap Europe 3 (STRATEGO): Translating the Heat Roadmap Europe Methodology to Member State Level. Available online: https://heatroadmap.eu/sp_faq/heat-roadmap-europe-3-stratego-2015/ (accessed on 30 April 2022).
[5] Corradi E, Rossi M, Mugnini A, Nadeem A, Comodi G, Arteconi A, et al. Energy, environmental, and economic analyses of a district heating (DH) Network from both thermal plant and end-users' prospective: an Italian case study. Energies 2021;14:7783. https://doi.org/10.3390/en14227783.
[6] Yuan M, Thellufsen JZ, Sorknæs P, Lund H, Liang Y. District heating in 100% renewable energy systems: Combining industrial excess heat and heat pumps. Energy Convers Manage 2021;244:114527. https://doi.org/10.1016/j.enconman.2021.114527.
[7] Van Oevelen T, Scapino L, Koussa J, al Vanhoudt D. A case study on using district heating network flexibility for thermal load shifting. Energy Rep 2021;7:1–8. https://doi.org/10.1016/j.egyr.2021.09.061.
[8] Vandermeulen A, van der Heijde B, Helsen L. Controlling district heating and cooling networks to unlock flexibility: a review. Energy 2018;151:103–15. https://doi.org/10.1016/j.energy.2018.03.034.
[9] Zhang Y, Johansson P, Kalagasidis AS. Applicability of thermal energy storage in future low-temperature district heating systems – Case study using multi-scenario analysis. Energy Convers Manage 2021;244:114518. https://doi.org/10.1016/j.enconman.2021.114518.
[10] Laakkonen L, Korpela T, Kaivosoja J, Vilkko M, Majanne Y, Nurmoranta M. Predictive supply temperature optimization of district heating networks using delay distributions. Energy Procedia 2017;116:297–309. https://doi.org/10.1016/j.egypro.2017.05.076.
[11] Bavière R, Vallée M. Optimal temperature control of large scale district heating networks. Energy Procedia 2018;149:69–78. https://doi.org/10.1016/j.egypro.2018.08.170.
[12] Lyons B, O'Dwyer E, Shah N. Model reduction for Model Predictive Control of district and communal heating systems within cooperative energy systems. Energy 2020;197:117178. https://doi.org/10.1016/j.energy.2020.117178.
[13] Drgoňa J, Arroyo J, Cupeiro FI, Blum D, Arendt K, Kim D, et al. All you need to know about model predictive control for buildings. Ann Rev Control 2020;50:190–232. https://doi.org/10.1016/j.arcontrol.2020.09.001.
[14] Verrilli F, Srinivasan S, Gambino G, Canelli M, Himanka M, Del Vecchio C, et al. Model predictive control-based optimal operations of district heating system with thermal energy storage and flexible loads. EEE Trans Autom Sci Eng 2017;14:547–57. https://doi.org/10.1109/TASE.2016.2618948.
[15] Hering D, Cansev ME, Tamassia E, Xhonneux A, Müller D. Temperature control of a low-temperature district heating network with Model Predictive Control and Mixed-Integer Quadratically Constrained Programming. Energy 2021;224:120140. https://doi.org/10.1016/j.energy.2021.120140.
[16] Saloux E, Candanedo JA. Model-based predictive control to minimize primary energy use in a solar district heating system with seasonal thermal energy storage. Appl Energy 2021;291:116840. https://doi.org/10.1016/j.apenergy.2021.116840.
[17] Vivian J., Jobard X., Hassine B.I., Hurink J.& Pietruschka D. Smart Control of a District Heating Network with High Share of Low Temperature Waste Heat. Conference: 12th Conference on Sustainable Development of Energy, Water and Environmental Systems - SDEWES 2017At: Dubrovnik (Croatia).
[18] Alanne K, Sierla S. An overview of machine learning applications for smart buildings. Sustainable Cities Soc 2022;76:103445. https://doi.org/10.1016/j.scs.2021.103445.
[19] Zhang L, Wen J, Li Y, Chen J, Ye Y, Fu F, et al. A review of machine learning in building load prediction. Appl Energy 2021;285:116452. https://doi.org/10.1016/j.apenergy.2021.116452.
[20] Sutton RBA. Reinforcement Learning: An Introduction. MIT Press; 1998.
[21] Wang Z, Hong T. Reinforcement learning for building controls: The opportunities and challenges. Appl Energy 2020;269:115036. https://doi.org/10.1016/j.apenergy.2020.115036.
[22] Fu Q, Han Z, Chen J, Lu Y, Wu H, Wang Y. Applications of reinforcement learning for building energy efficiency control: A review. J Build Eng 2022;50:104165. https://doi.org/10.1016/j.jobe.2022.104165.
[23] Azuatalam D, Lee WL, de Nijs F, Liebman A. Reinforcement learning for whole-building HVAC control and demand response. Energy and AI 2020;2:100020. https://doi.org/10.1016/j.egyai.2020.100020.
[24] Zhang G, Hu W, Cao D, Zhang Z, Huang Q, Chen Z, et al. A multi-agent deep reinforcement learning approach enabled distributed energy management schedule for the coordinate control of multi-energy hub with gas, electricity, and freshwater. Energy Convers Manage 2022;255:115340. https://doi.org/10.1016/j.enconman.2022.115340.
[25] Sakkas NP, Abang R. Thermal load prediction of communal district heating systems by applying data-driven machine learning methods. Energy Rep 2022;8:1883–95. https://doi.org/10.1016/j.egyr.2021.12.082.
[26] Wei Z, Zhang T, Yue B, Ding Y, Xiao R, Wang R, et al. Prediction of residential district heating load based on machine learning: A case study. Energy 2021;231:120950. https://doi.org/10.1016/j.energy.2021.120950.
[27] Geysen D, De Somer O, Johansson C, Brage J, Vanhoudt D. Operational thermal load forecasting in district heating networks using machine learning and expert advice. Energy Build 2018;162:144–53. https://doi.org/10.1016/j.enbuild.2017.12.042.
[28] Maljkovic D, Basic BD. Determination of influential parameters for heat consumption in district heating systems using machine learning. Energy 2020;201:117585. https://doi.org/10.1016/j.energy.2020.117585.
[29] Solinas FM, Bottaccioli L, Guelpa E, Verda V, Patti E. Peak shaving in district heating exploiting reinforcement learning and agent-based modelling. Eng Appl Artif Intell 2021;102:104235. https://doi.org/10.1016/j.engappai.2021.104235.
[30] Brandi S, Fiorentini M, Capozzoli A. Comparison of online and offline deep reinforcement learning with model predictive control for thermal energy

management. Autom Constr 2022;135:104128. https://doi.org/10.1016/j.autcon.2022.104128.
[31] Ceusters G, Rodríguez RC, García AB, Franke R, Deconinck G, Helsen L, et al. Model-predictive control and reinforcement learning in multi-energy system case studies. Appl Energy 2021;303:117634. https://doi.org/10.1016/j.apenergy.2021.117634.
[32] Comodi G, Lorenzetti M, Salvi D, Arteconi A. Criticalities of district heating in Southern Europe: Lesson learned from a CHP-DH in Central Italy. Appl Therm Eng 2017;112:649–59. https://doi.org/10.1016/j.applthermaleng.2016.09.149.
[33] Mugnini A, Comodi G, Salvi D, Arteconi A. Energy flexible CHP-DHN systems: Unlocking the flexibility in a real plant. Energy Convers Manage: X 2021;12:100110. https://doi.org/10.1016/j.ecmx.2021.100110.
[34] Serale G, Fiorentini M, Capozzoli A, Bernardini D, Bemporad A. Model Predictive Control (MPC) for Enhancing Building and HVAC System Energy Efficiency: Problem Formulation. Appl Opport Energies 2018;11:631. https://doi.org/10.3390/en11030631.
[35] Rawlings JB; MDQ. Model Predictive Control: Theory and Design. 2012.
[36] Ministero dello Sviluppo Economico (MISE). Decreto Ministeriale 26 Giugno 2015. (In Italian) Applicazione delle metodologie di calcolo delle prestazioni energetiche e definizione delle prescrizioni e dei requisiti minimi degli edifici. Available online: Gazzetta Ufficiale. 2015. Available online: www.gazzettaufficiale.it/eli/id/2015/07/15/15A05198/sg (accessed on 30 April 2022).
[37] Agenzia Nazionale per le Nuove Tecnologie (In Italian), L'energia e lo Sviluppo Economico Sostenibile (ENEA). (In Italian) Rapporto Energia e Ambiente 2002. Available online: old.enea.it/com/web/pubblicazioni/REA_02/analisi_02.pdf (accessed on 30 April 2022).