SimBuild
2024

Eleventh National Conference of IBPSA-USA
Denver, Colorado
May 21 – 23, 2024

Optimizing Operational Costs in Combined Heat and Power Integrated District Heating Systems: A Reinforcement Learning Approach

Saranya Anbarasu[1], Tanmay Ambadkar[1], Rosina Adhikari[1],
Kathryn Hinkelman[1], Zhanwei He[1], Wangda Zuo[1,2], Ardeshir Moftakhari [1]
[1]Pennsylvania State University, State College, PA
[2]National Renewable Energy Laboratory (NREL), Golden, CO

Abstract

As societies worldwide strive to reduce carbon footprints and transition toward cleaner energy sources, grid-integrated district energy systems (DES) emerge as a pivotal player. Meanwhile, the escalating complexity of DES necessitates adaptive, synergistic, and hierarchical control of heterogeneous systems to achieve common energy and cost conservation goals. Prior research highlights several challenges of model-based control techniques for DES, such as limited access to computational tools, prolonged durations to digital twin development, and the complexities associated with control design. In contrast, model-free control methodologies may be a viable alternative. As a response, this study explores a reinforcement learning-based (RL) supervisory control to minimize the operational costs in a university campus DES. To enhance overall system efficiency, we utilize resource flexibility to improve DES operations by responding to fluctuations in utility prices. In this paper, we demonstrate the toolchain, develop the virtual testbed, engineer a suitable RL reward, and learn from challenges. From the case study, the RL agent showcases a significant 32% net operational cost savings and a 13% peak demand reduction compared to the conventional thermal load following control. This research signifies the potential of RL-based control systems in optimizing the performance of complex DES and multi-energy systems.

Introduction

The global imperative to address climate change has led to a significant rise in the adoption of district energy systems (DES), which improved grid interactivity further. These centralized systems typically operate by generating steam, hot water, chilled water, and electricity in a central plant. These utilities are then distributed through an insulated network of pipes and electric cables to meet the thermal and electrical needs of interconnected buildings. Particularly in densely populated urban areas, these systems play a crucial role in enhancing sustainability and providing cost-effective solutions to achieve decarbonization goals (Allegrini et al. 2015; Lund, Ilic, and Trygg 2016; Bataille et al. 2016). Modern DES integrate more energy-efficient technologies, including advanced heat pump systems, ambient loop heating and cooling mechanisms, waste heat recovery initiatives, combined heat and power systems, and various renewable energy sources and storage options (Revesz et al. 2020). Additionally, these systems contribute to the resilience and adaptability of the grid by offering essential support services such as demand response, frequency regulations, and other ancillary functions.

Despite the recognized benefits of DES, a significant portion of studies in this domain remain predominantly theoretical, lacking comprehensive modeling platforms (Allegrini et al. 2015; Klemm and Vennemann 2021), thermodynamic validation, and tangible real-world applications (Heendeniya, Sumper, and Eicker 2020). While the industry benefits from steady-state calculators and spreadsheets for techno-economic and environmental projections, there is a noticeable gap in tools capable of dynamically predicting and controlling the performance of DES under various scenarios including uncertainties (Allegrini et al. 2015). There are a variety of control methods that have been used extensively in the past, ranging from manual operator-driven decisions, rule-based controls (RBC), to advanced model-predictive controls (MPC) integrated into SCADA systems. However, the field of model-free controls is emerging as a promising option that addresses the limitations of previous methods. RBCs face challenges in defining rules for new system additions, lacking anticipation and optimization capabilities over specific time horizons (Saloux and Candanedo 2020). While, MPCs require high fidelity calibrated models (Verrilli et al. 2016), fast computational simulation speed, susceptible to demand forecast errors (Verrilli et al. 2016), and results in system safety violations in real-world applications (Hermansen et al. 2022).

Unlike MPC, RL is a model-free machine learning technique that does not require high-fidelity models for control decisions and can learn directly from the environment (Sutton and Barto 2018). RL algorithms, such as deep Q learning, improve thermal energy efficiency in district heating consumer facilities (Testasecca et al. 2023). Integration of optimization and deep learning RL modules shows promise for the fast online optimization of district heating systems (Lee et al. 2020). RL applied to large-scale multi-source looped district cooling systems can save up to 12.99% of annual distribution energy consumption through Modelica-python co-simulation (Wang et al. 2023). Comparing MPC and RL, for a hydronic heat pump system integrated with BOPTEST emulators for building models (Blum et al. 2021), reveals that RL controllers using the deep deterministic policy gradient algorithm outperform the baseline controller in both typical and peak heating scenarios, while outperforming MPC under uncertainties. On- and off-policy multi-objective RL agents can even outperform perfect foresight Linear MPC in multi-energy system configurations while achieving stabilization in the reward after 5.7 years (200,000 training time-steps) (Ceusters et al. 2021). Therefore, drawing on past experiences, RL agents fine-tunes control actions and proves adept at managing complex energy systems. However, trade-offs such as extended training periods, imbalances from random explorations, and the delicate balance between swift responses and system safety exist (Wang et al. 2023).

Responding to model-based control challenges and the need for new toolchains, we explore minimum operational cost optimization using RL-based control with a Modelica virtual DES testbed. By doing so, this work addresses three critical gaps in the literature. First, we address time and computational constrain by leveraging the model-free control method for complex system control application. Second, we demonstrate the use of the dynamic Modelica virtual DES testbed for training the RL agent, which is less explored. Specifically, the use of dynamic models ensures system stability, which are the limitations of linear models as they cannot predict critical system process outputs (eg., steam outlet pressure and steam drum void fractions in steam boilers, condensation in boiler economizers which can derate the component, etc). Third, we reduce the extended training period in the dynamic environment through control disintegration and RL reward engineering, and our learning have been reported in the discussion section.

In the subsequent sections, we begin by outlining the methodology, followed by the discussion of dynamic system models and the integrated virtual testbed. Subsequently, we describe the problem formulation, simulations, and results discussions.

Methodology

This paper demonstrates and evaluates RL-based DES controls for a university campus DES. The following sections outline the steps to accomplish this, from understanding the system and establishing the calibrated virtual testbed, to coupling the RL-Modelica tool and running the simulations.

Understanding The System

The case study site for this work is the district heating and power systems at the University of Colorado Boulder campus. Our initial step involves understanding the campus DES and baseline controls by collecting system descriptions, sizing information, and operational data from campus monitoring systems. Figure 1 illus-

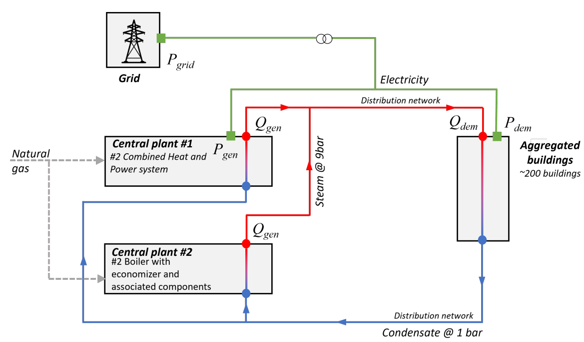


Figure 1: Schematic of the district energy system

trates the schematic of the university campus DES, located in ASHRAE Climate Zone 5b. The system comprises two central plants. The first houses a district heating system equipped with 22,679 kg/hr and 45,359 kg/hr dual fuel boilers. These boilers are equipped with flue gas to feed water economizers and an associated feed water pumping and condensate return system. The second plant accommodates two 15MW Combined Heat and Power (CHP) systems with additional firing capabilities, incorporating duct burners into the Heat Recovery Steam Generator (HRSG). Both plants supply saturated steam at 9.0 bar to campus buildings through underground insulated pipelines within distribution tunnels. Due to the unavailability of heating and electricity demand time series profiles of individual buildings and to avoid computational stress of intricate distribution network configurations, the steam and electricity demands of the buildings are aggregated. Further, the buildings are grid-tied, enabling the utilization of power beyond that generated by the CHP systems. Regarding controls, the university campus DES follows boiler auto control and manual biasing override by operators. Lastly, because the CHP system currently serves as a peaker plant to the utility, this case study considers a thermal load following control (TLF) strategy as baseline.

Modelica Testbed Implementation

We use Modelica, an equation-based object-oriented language, to develop the high-fidelity virtual testbed model of the campus DES. Moreover, Modelica has been increasingly used for building and district energy simulation, mainly due to the capabilities of flexible and hierarchical modeling of systems, and controls, integrated with its thermodynamic and hydraulic performances (Wetter and van Treeck 2017; Hinkelman et al. 2022c; Müller et al. 2016). Based on the acquired system information, we develop an integrated virtual testbed, leveraging component models from the Modelica Buildings Library (MBL) (Wetter et al. 2014), specifically MBL v10.0.0. The MBL includes a wide range of fluid and electric component models, control blocks, and an exclusive package featuring models for DES. Additionally, we implement specific system models, including a performance curve-based CHP model and baseline control modules developed for this case study. Detailed Modelica model descriptions can be found in the successive section: *DES Testbed.*

For this study, we first hierarchically calibrate and validate individual sub-systems. Then, we integrate the subsystems into a comprehensive virtual testbed of the campus. Given our emphasis on facilitating multiple simulation iterations to train the RL agent, we strategically simplify the system by removing certain components with large non-linear functions (such as control valves, dynamic switches, etc). This simplification is carefully executed to balance the need for computation speed without sacrificing model accuracy and system dynamics. The results of virtual testbed calibration and validation agreed with ASHRAE Guideline 14 standards, with CVRMSE $\leq 11\%$ and MBE $\leq 6.8\%$ for evaluated parameters, such as the natural gas consumption, power generated by CHP, boiler exhaust temperature, boiler outlet steam pressure, and economizer inlet/outlet feed water temperatures.

FMU Export

The functional Mock-up Unit (FMU) standard enables the tool-independent exchange of dynamic models in binary format (Blockwitz et al. 2012). In the context of a Modelica model, an FMU refers to a file that encapsulates the dynamic system model, allowing it to be easily shared and integrated into various simulation environments that support FMUs. To facilitate control of the DES by the RL agent, we exposed the supervisory control parameters in the Modelica model. The FMU is exported using the *Optimica Toolkit Compiler* (Åkesson et al. 2010), which is built using pyFMI, enabling external interactions with FMUs through Python.

Reinforcement Learning

The RL agent, represented by a machine learning algorithm, takes actions in the environment (DES virtual testbed), and receives rewards or penalties as feedback. The objective of the agent is to continuously revise its strategy over time to maximize cumulative rewards.

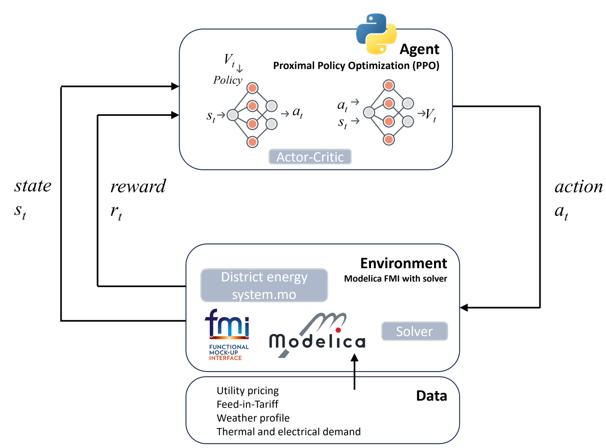


Figure 2: Modelica-RL toolchain and methodology

In brief, RL is broadly categorized into *value-based methods* (e.g., Q-learning, Deep Q Network), estimating expected rewards for actions, and *policy-based methods* (e.g., deep deterministic policy gradient, actor-critic, proximal policy optimization), directly determining optimal decision-making policies. For this work, we use proximal policy optimization (PPO), a modification of the actor-critic framework that addresses challenges related to stability and sample efficiency in training policies using Kullback-Leibler divergence (Schulman et al. 2017). The actor-critic paradigm merges aspects of both value-based and policy-based methods, leveraging the strengths of each. The objective function of PPO is

$$L(\theta) = \mathbb{E}_t \left[\min \left(\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)} A_t, \right. \right.$$

$$\left. \left. \text{clip} \left(\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)}, 1-\varepsilon, 1+\varepsilon \right) A_t \right) \right], \quad (1)$$

where $L(\theta)$ is the surrogate objective; $\pi_\theta(a_t \mid s_t)$ is the probability of taking action a_t in state s_t under the current policy; $\pi_{\theta_{\text{old}}}(a_t \mid s_t)$ is the probability of taking action a_t in state s_t under the previous policy; A_t is the advantage function representing the difference between the estimated value of the state s_t and the sum of rewards; and ε is a hyperparameter controlling the size of the trust region. The value function of the PPO is updated using the mean squared error loss to minimize the difference between the predicted and the actual values, given as

$$L_{\text{VF}}(\phi) = \mathbb{E}_t \left[\left(V_\phi(s_t) - \hat{V}_t \right)^2 \right], \quad (2)$$

where $L_{\text{VF}}(\phi)$ is the value function loss; $V_\phi(s_t)$ is the predicted value of the state s_t; and $\hat{V}_t$ is the estimated value of the state s_t, which is usually the sum of the rewards.

Problem Formulation

The RL objective is to minimize the operational costs of the DES by responding to various grid signals, such as natural gas prices (C_g), Time-of-Use (TOU) electricity prices (C_i), and Feed-In Tariff (FIT) (C_e) for selling excess electric power back to the grid. In this work, the agent dynamically adjusts the firing bias ratio of the boiler, CHP, and HRSG duct burners to minimize the overall operational cost, while ensuring high operational efficiency. The bias or capacity ratios are the supervisory control parameters, which are used to shift load between multiple boilers/CHP systems. For example, 15% negative bias ensures the boiler firing capacity is limited to 85%. Constraints include thermal and electricity demand balance, the threshold of firing signals (y_{max} and y_{min}), and output steam pressure setpoint p_{set}. The implementation involves finite-dimensional state and action spaces denoted as $S \subset \mathbb{R}$ and $A \subset \mathbb{R}$, respectively. The *action space* of the agent includes the control signal sent to the environment, which includes the bias ratios, represented as

$$A \subset \{yB_{b1}, yB_{b2}, yB_{ctg1}, yB_{ctg2}, yB_{hrsg1}, yB_{hrsg2}\} \in \{0,1\}, \quad (3)$$

where yB_n are the system bias ratios, with subscripts defining each DES subsystem. The *state space* is defined as a combination of controllable, uncontrollable, and temporal components. The *controllable* is the observable state of the model that has a direct relationship with the actions, given as

$$S \subset \{p_{b1}, p_{b2}, p_{hrsg1}, p_{hrsg2}, E_{ctg1}, E_{ctg2}, E_{grid}, \sum \dot{m}_{g,n}\}, \quad (4)$$

where p_b and p_{hrsg} are the boiler and HRSG steam header pressures (Pa); E_{ctg} is the electricity generated by the CHP (kWh); and $\sum \dot{m}_{g,n}$ is the total fuel consumption of the gas turbine, HRSG, and boilers (kg). The *uncontrollable* state spaces consist of states that are independent of the agent's actions, given as

$$S \subset \{C_i, C_e, C_g, T_{dryBul}, \dot{Q}_{dem}, E_{dem}\} \quad (5)$$

where $\dot{Q}_{dem}$ and E_{dem} are the steam (kg/s) and electricity demand signals (kWh) from aggregated buildings and T_{dryBul} is the outdoor drybulb temperature (K). The temporal aspect is encapsulated within the state space as $S \subset \{t\}$, which represents the simulation time information. In RL problems, constraints are incorporated along with the reward function, and emphasis is given through the inclusion of weightage or scaling factors. The reward function in any RL algorithm is inherently a maximization function. However, our goal is to minimize costs. As a solution, we indirectly formulated the reward functions by utilizing the complement of normalized values for net energy cost and steam header pressure. To assign significance to each reward component, linear and quadratic functions were incorporated, defined as

$$r_t = S_1 \overbrace{\frac{1}{(1 - p_{norm,n,t})}}^{\text{pressure reward}} + \overbrace{S_2(1 - (E_i C_i - E_e C_e)_{norm,t})^2}^{\text{energy cost reward}}$$

$$+ S_3 \underbrace{\frac{1}{(\sum \dot{m}_{g,n} C_g)_{norm,t}}}_{\text{gas cost reward}}, \quad (6)$$

where E and $\dot{m}_g$ are the electricity (kWh) and natural gas consumption (kg/s); C is cost/unit; n denotes the DES subsystems; subscripts i and e are the electricity import and export; dem is the building demand; and ctg, $hrsg$, and boi represent the combustion gas turbine, HRSG, and boiler, respectively; $S_1, S_2 and S_3$ represent weightage factors. To note, this study does not currently factor in demand charges for RL reward estimation, which are associated with peak power consumption.

Co-simulation Toolchain

Utilizing the Python package *stable-baselines 2.10.2* (Raffin et al. 2019), we implement the RL agent. Figure 2 illustrates the RL process that integrates the FMU of the DES environment with the agent. The FMU is simulated with *CVode solver* (Cohen, Hindmarsh, and Dubois 1996), and external integration handles additional required data for simulating the environment, such as weather profiles, utility rates, aggregated demand signals, and other information provided to the agent. At each time iteration t, the states s_t and the reward r_t from the environment are communicated to the agent. Using PPO, the agent determines the action a_t which is returned to the environment. Initially, the agent takes random exploration, and the policy is eventually refined after undergoing extensive training.

Simulation Settings

We trained the RL agent to suggest actions within 6 action spaces (A) based on 22 observational spaces (S) over 31 days (January 2022), utilizing campus demand signals, electric utility prices, and the weather data for Broomfield, CO, USA. The agent interacts with the environment every $t = 900$s. Since the annual DES demand fluctuates with large variance, we considered training the RL with a month of data for this initial exploratory work. Limiting the variance of demand profile enables us to understand the toolchain challenges, reward engineering process, and the benefits of various training periods. In the future, we plan to train the agent for multiple years of utility prices, demands, and weather profiles. The settings used for training the RL agent are shown in Table 1.

Table 1: RL agent settings

Parameter	**Value**
RL policy	PPO-Mlp policy
n_steps	64
n_epochs	30
$learning\_rate$	0.00005
Time steps trained	10^5

For every iteration t, all $A_t \subset \mathbb{R}$ are used as inputs by the FMU (environment) to simulate the model. After training 10^5 steps, we tested the RL agent-based control and compared the performance with the baseline control. The parameters considered for this comparison include the system pressures (p_{b1}, p_{b2}, p_{hrsg1}, and p_{hrsg2}), fuel consumption ($\sum \dot{m}_{g,n}$), electricity generation (E_{dem}), and net operational cost ($E_iC_i - E_eC_e$).

DES Testbed

We implement the virtual testbed as integrated subsystems, including the district heating plant, the CHP powerhouse, aggregated building heating and electrical loads, and the corresponding controls. The building heating loads are measured at the energy transfer stations, the electric loads are modeled using a resistance-capacitance model. To develop the virtual testbed we leverage Modelica models from MBL v10.0.0 (Wetter et al. 2014) and incorporate models from our previous DES research (Hinkelman et al. 2022a; Hinkelman et al. 2022b).

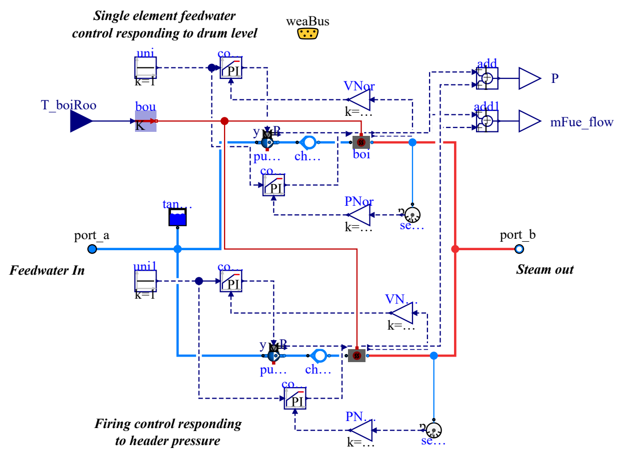
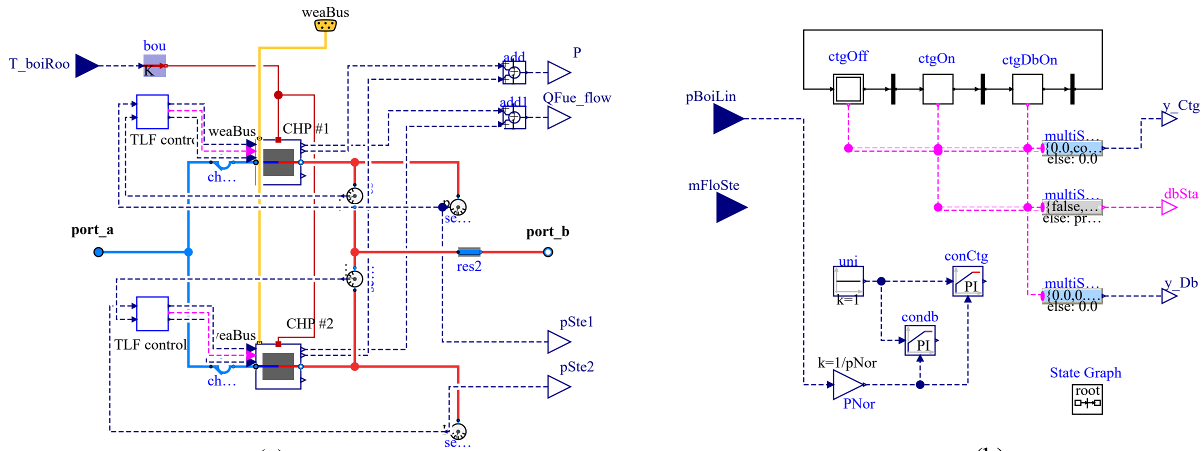


Figure 3: Modelica diagram of the district heating plant

District Heating Plant Model

In Figure 3, the Modelica model illustrates the district heating plant with two boilers integrated with feedwater pumps, an expansion tank, check valves, and associated control systems. The `port_a` represented the feedwater inlet and `port_b` represents the steam outlet from the plant. An improved steam medium model, specifically `Buildings.Media.Steam`, is used for its computational efficiency and accuracy (Hinkelman et al. 2022b). The boiler uses a split-medium approach that efficiently models the water/steam phase change and the two-phase region separately (`Buildings.Experimental.DHC.Plants.Steam.BaseClasses.BoilerPolynomial`). This approach serves to decouple the non-linearities associated with phase change states, resulting in an overall improvement of simulation speed. The boiler model incurs radiation and convection losses based on the boiler's room temperature (`T_boiRoo`). Pressure sensors are positioned at the boiler outlet port, which monitors the boiler header pressure; the boiler firing rates are adjusted through a PI controller to maintain the steam pressure setpoint of 9 bar. Correspondingly, the steam drum level is measured using the volume of water within the control volume



(a) (b)

Figure 4: Modelica model of the powerhouse, component, and control

of the boiler model. To maintain the drum level at 0.5 of the drum volume, a PI controller modulates the feed water pump. The heating plant model outputs the auxiliary power consumption of the feed water pumps (`P`) and the mass flow rate of fuel (`mFue_flow`).

CHP Powerhouse Model

The CHP power house consits of two CHP systems integrated with TLF controls, as represented in Figure 4. We implement a new performance curve-based CHP model to support quick simulations when integrated with the RL agent (see Figure 5). The efficiencies of CTG η_{ctg} and the HRSG with/without auxiliary duct burner firing η_{hrsg} are estimated using the performance curves. The η_{ctg} and η_{hrsg} vary for different part load conditions y_{ctg} and y_{hrsg}, inlet air temperature T_{dryBul}, and the exhaust temperature at the outlet of CTG T_{exh}, formulated as

$$\eta_{ctg} = f(y_{ctg}, T_{dryBul}) \quad \text{and} \tag{7}$$

$$\eta_{hrsg} = f(y_{hrsg}, T_{exh}), \tag{8}$$

where the performance curves are given as quadratic splines with 50 data points. The η_{ctg} is larger at lower T_{dryBul}, as the usable torque increases with a higher density of inlet air. The fuel consumed by the CTG and HRSG are

$$\dot{m}_{fue,ctg} = \frac{y_{ctg} P_{nom}}{\eta_{ctg}\, LHV_{fue}} \quad \text{and} \tag{9}$$

$$\dot{m}_{fue,hrsg} = \begin{cases} \dfrac{y_{ctg}(\dot{Q}_{nom} + \dot{Q}_{exh})}{\eta_{ctg}\, LHV_{fue}} & if\ y_{hrsg} > 0 \\ \dfrac{y_{ctg} \dot{Q}_{exh}}{\eta_{ctg}\, LHV_{fue}} & if\ y_{hrsg} = 0. \end{cases} \tag{10}$$

The usable heat $\dot{Q}_{wat}$ that is transferred to the feed water is modeled as a control volume (`Buildings.Experimental.DHC.Plants.Steam.BaseClasses.ControlVolumeEvaporation`) from MBL, and the heat loss from the HRSG shell $\dot{Q}_{loss}$ is modeled as a heat capacitance (`Modelica.Thermal.HeatTransfer.Components.HeatCapacitor`) and conduction component (`Modelica.Thermal.HeatTransfer.Components.ThermalConductor`) from the Modelica standard library, as represented in Figure 5. Mathematically this is

$$\begin{aligned} \dot{Q}_{wat} &= \eta_{hrsg}(y_{hrsg}\dot{Q}_{nom} + \dot{Q}_{exh})\ if\ y_{hrsg} > 0 \\ \dot{Q}_{wat} &= \eta_{hrsg}\dot{Q}_{exh}\ if\ y_{hrsg} = 0, \end{aligned} \tag{11}$$

where $\dot{Q}_{nom}$ and $\dot{Q}_{exh}$ denote the nominal heating capacity of the duct burners and recoverable heat in exhaust from the CTG in kWh, respectively; and LHV_{fue} denotes the low heating value of the fuel in kJ/kg.

Thermal Load Following Control (TLF)

Common control strategies employed in CHP systems include Thermal Load Following (TLF) and Electrical Load Following (EFL). In this investigation, TLF is implemented to dynamically match the campus steam demands in real-time by staging the CTG and HRSG duct burners in the CHP power house. The electrical power generated in this process is considered an additional benefit. Figure 4(b) represents the Modelica model of TLF control state graph along with PI controllers that first govern the staging of CHP and HRSG, and then determines the firing rates. The state transitions are fired based on the threshold of the maximum steam genera-

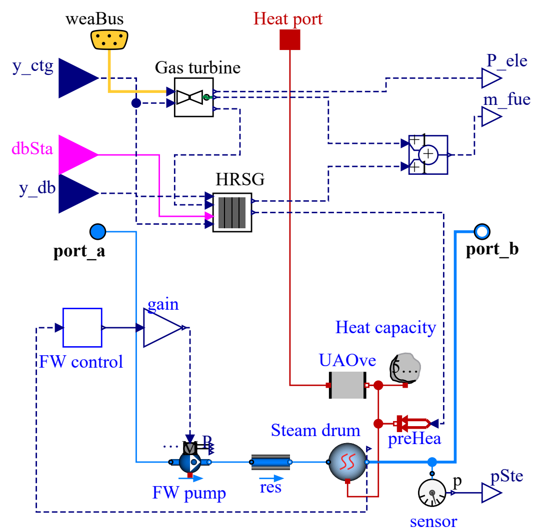


Figure 5: Modelica diagram of the Combined heat and power (CHP) model

tion capacity of each component.

Thermal and Electrical Demand Model

The building steam and electrical demands are aggregated due to a lack of individual building demand profiles and a focus on plant-level controls. The aggregated building load in ETS is modeled using components from MBL that function as a condenser volume, steam trap, and condensate return pump (see Figure 6). The PI controller modulates the condensate pump based on the steam demand input of the model. Similarly, the building electrical loads are aggregated, which includes the power consumption of buildings, chillers, cooling towers, combustion control fans, and pumps in the district plant (see Figure 6). The electric load is modeled as a resistance-capacitance load with a static power factor of 0.8. This load is connected to the CHP and the power grid (3-phase balanced, 480V, `Buildings.Electrical.AC.ThreePhasesBalanced.Sources.Grid`).

We calibrate and validate each of the discussed system models for optimal performance in comparison to the measured data from the university campus DES. The virtual testbed model in Figure 6 integrates the baseline controls, distribution pipes, campus steam, and electricity demand. For the RL environment, we revise the testbed model to include the bias factors as input signals. These supervisory control parameters are modified by the agent during simulation in response to rewards. At local levels, each of the components and subsystems has individual firing and feed water controllers.

Results

After the successful training period of the RL agent, we executed the simulation and compared the RL agent's performance to the baseline control. In this section, we first discuss and visualize the control policy. Then, we present the operational cost-saving potential of RL relative to the baseline case.

Control Policy Visualization

Steam Header Pressure

The primary constraint of the optimization problem is to maintain the steam pressure setpoint in the model. The steam header pressure drops when the demand increases and the local PI controllers fire the boiler and HRSG systems to modulate the pressure back to the set point. Figure 7 presents the steam header pressure measured at the outlet of boilers and HRSGs. Both the baseline and RL control maintain the operating pressure of the steam systems (9 bar). This indicates that the supervisory control parameters (i.e., the bias ratios output by the RL agent) are ensuring system safety and stability.

Operational Efficiency

Figure 8 presents the operational efficiencies of the boiler, CHP, and HRSG for a 7-day period. The boiler efficiencies are consistently 70-78% for both baseline and RL cases. The operational efficiencies of the CHP and HRSG are highly dynamic due to the outdoor air temperature and part-load factors. Since the RL case maintains consistent firing unlike the baseline case, the electrical efficiency η_{chp} is stable at 30%, while the baseline case has a lower η_{chp} under low load conditions (60-72 hours). The incremental firing of duct burners in HRSG results in a higher thermal efficiency $\eta_{hrsg} \approx 90\%$, which is utilized by the RL agent to save energy cost (also reflected in Figure 10).

Electricity Imports, Exports, and Generation

Figure 9 illustrates the outcomes of electricity generation by the CHPs and the imports/exports of electricity from/to the grid. We observe that the RL case generates higher electricity compared to the baseline case for all hours, yet there is a further difference during the high TOU periods (see Figure 9(a)). The RL agent increases the firing bias factors of CHP and HRSG during high TOU periods, ensuring lower import from the grid. This for example is observed between 60-72 hours and is characterized as a load-shedding demand response action. During low TOU periods (e.g., 24-36 hours),

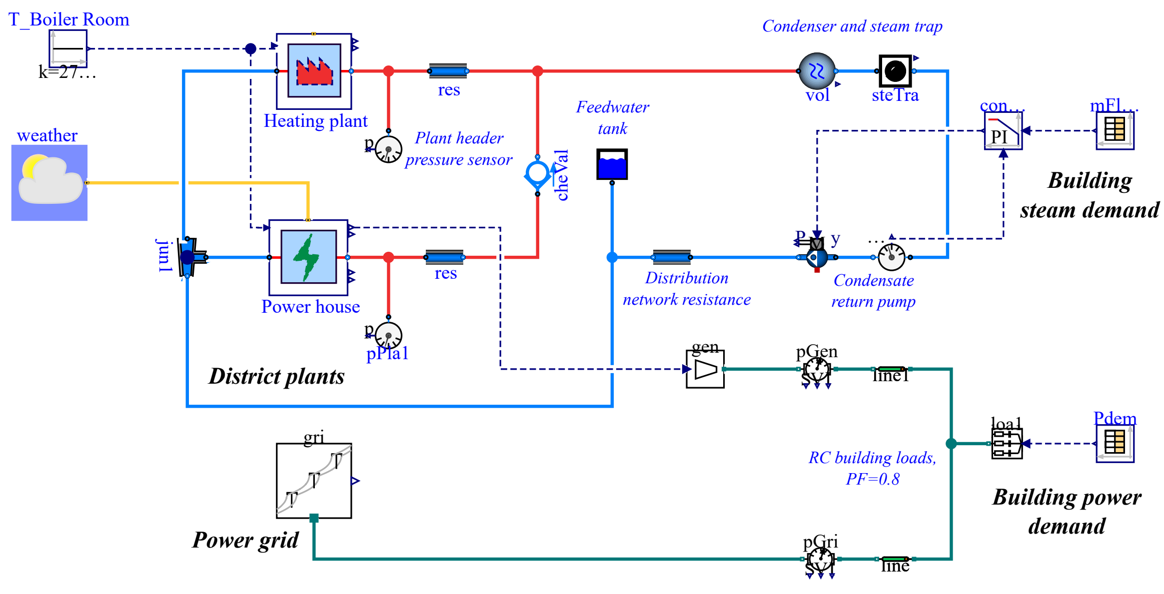


Figure 6: Modelica model of case study system

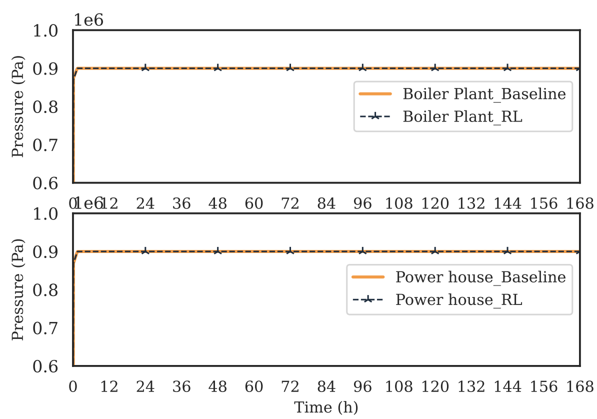


Figure 7: Steam header pressure comparison between baseline controls and RL agent-based controls

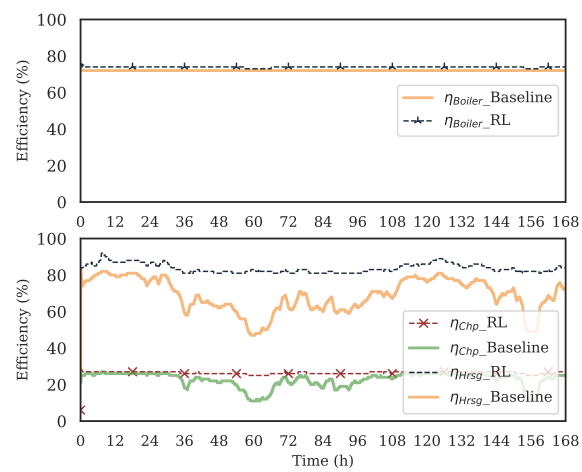


Figure 8: Operational efficiency of boilers and CHP

the RL agent performance closely aligns with the baseline control, yet exhibits a slight increase in generation and export to the grid (see Figure 9(b)). The prices of natural gas and the FIT remain constant over time, resulting in no variation in RL decisions in response to these prices. Although we observe load shedding, there is no significant load shifting observed in either case, since the system's thermal generation is constrained by the campus steam demand and the unavailability of thermal/electrical storage.

Cost-Saving Potential

The net electricity cost is calculated based on the electricity imports and exports from the grid based on the TOU and FIT utility rates. The natural gas and FIT prices are constant at \$0.050/kg/s and \$0.045/kWh, respectively, while the TOU electricity buying price varies as represented in Figure 9(d). The operational costs for electricity and natural gas, with low and high TOU periods highlighted in Figure 10. The cost of electricity for the baseline is higher than the RL case, which

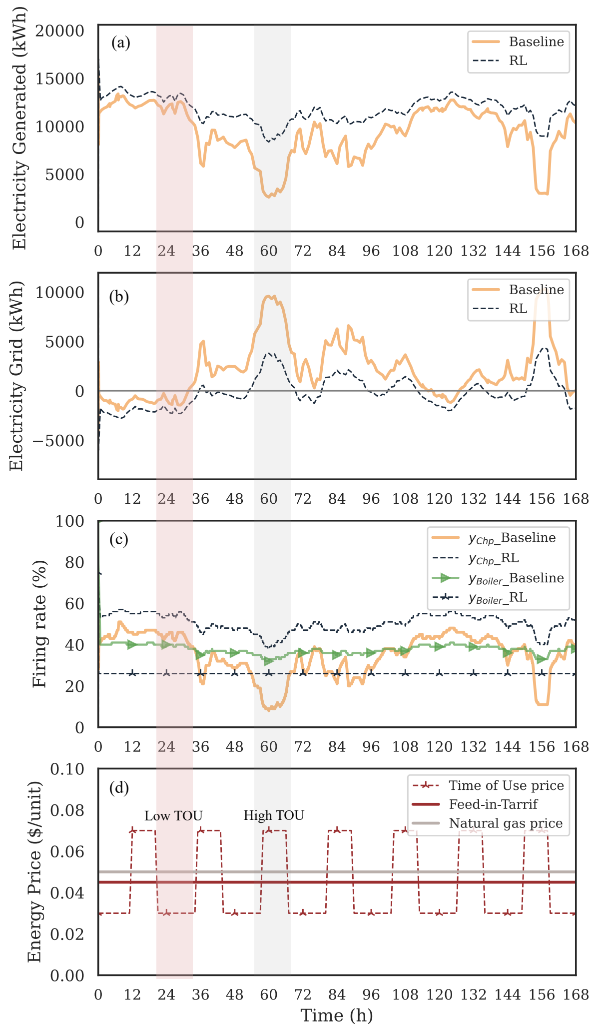


Figure 9: Comparison of electricity imported and generated in response to utility pricing

peaks due to excess imports during high TOU periods (60-72h). This agrees with the energy from grid trend in Figure 9(b). It is interesting to note that the magnitude of energy peaks and cost peaks are different due to TOU. However, the exported electricity (observed as negative energy costs) has a similar profile for baseline and RL cases. On the other hand, the natural gas cost of the RL case is higher than the baseline case, reflecting the excess energy generation beyond the campus electric demand. This is primarily attributed to the increased part-load operation of the CHP, which meets both thermal and electrical demands while keeping the boilers on baseload. (refer to Figure 9(c)). Since the price of natural gas is lower than electricity (Figure 9(d)), excess onsite generation in the RL case at higher operational efficiencies is more feasible.

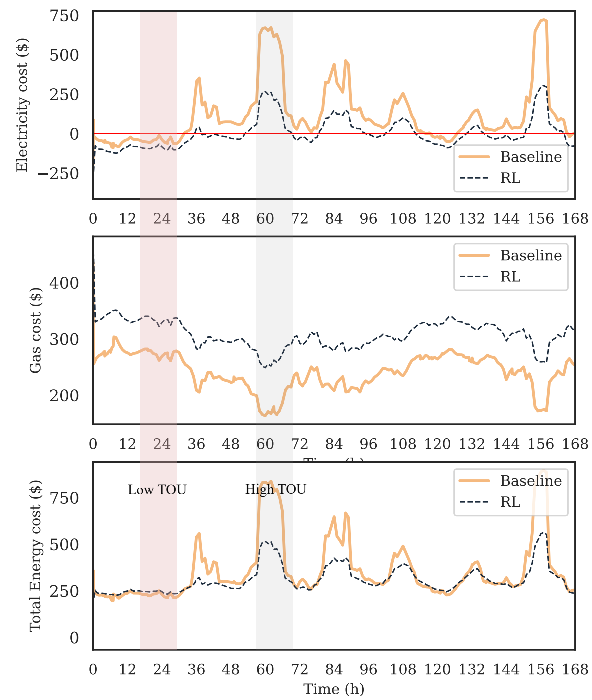


Figure 10: Comparison of electricity and natural gas consumption cost for one week

Table 2 presents the summary of results for the 31-day simulation period in January 2022. We observe the total electricity generation of RL case to be 15.99% higher than the baseline, while the imported electricity from the grid is -123% lower, reflecting the RL's decision to sell excess generation back to the grid. This is also reflected in the electricity cost, where the RL case is in credit, demonstrating a net-positive electricity case. However, the natural gas fuel cost is 30.47% higher for the RL case, reflecting the excess onsite electricity generation. Moreover, the net energy cost including both natural gas and electricity is found to be 32.30% lower for the RL case compared to the baseline. In addition, beyond the energy consumption, the peak load of the campus (import maximum) is reduced by 13.20% for the RL case, which can have cost savings in electricity demand charges.

Discussion

Through this work, we demonstrate the application of RL-based integrated system control for grid-integrated DES. The results show promising energy cost and peak

Table 2: Summary of results for 31 days of testing

Parameter	Unit	Baseline	RL case	Difference
Energy generated	kWh	3,975,070	4,610,661	15.99%
Energy from the grid	kWh	190,964	-44,621	-123.37%
Peak power	kWh	10,995	9,544	-13.20%
Gas consumed	kg	6,242	8,144	30.47%
Electricity cost	$	202,575	-3,907	-101.93%
Gas cost	$	224,736	293,197	30.39%
Net cost	$	427,311	289,290	-32.30%

demand savings of 32.3% and 13.2% respectively. By interacting with the virtual testbed environment, the RL agent determines optimal control actions corresponding to six different systems, highlighting its feasibility of handling complex systems with multiple control points. Moreover, unlike the model-based controls (i.e., MPC), this model-free approach could save time in developing high-fidelity MPCs via data-driven or reduced-order models. With a dynamic Modelica model that reflects the real environment, this work precisely observed the system's thermodynamic and engineering violations due to agent actions. While this work is under exploratory stages, progressive training of the agent for various campus demands and weather conditions can exhibit the adaptiveness of this method.

This work illuminated important challenges and insights for RL-based control of grid-integrated DES. Regarding the *dynamics of the environment*, managing complex thermodynamic processes embedded in the deeper layers of Modelica models becomes particularly challenging when these models interact with dynamic fluid media such as steam. The sensitivity of these models to inputs is heightened, leading to potential model failures when the agent tests random values. For example, the firing rates controlled by the agent directly influence the steam outlet pressure of the plants. Hence, we disintegrated the controls hierarchically based on the control time scales and included supervisory control parameters that were determined by the RL agent. This disintegration resulted in a series of local controls that dynamically control the processes, ensuring system stability, yet supervised by bias factors to shift/modulate the steam and electricity generation. Unlike linear or reduced-order models, integrating a high-fidelity model with RL agents can provide better insights into system failures and threshold violations.

Further, insights were also gained regarding the *reward function formulation*. The trajectory of RL learning is determined by the reward function, action, and observational spaces. As such, it is crucial to carefully select observational variables that effectively characterize the entire system. For example, we found that minimizing the observation space by removing a few (un)controllable states such as feedwater pump speed, duct burner status, and boiler room temperature did not support RL learning rate. The agent is primarily a neural network model, where the addition of unnecessary variable increases the complexity of the network, yet does not not contribute to enhanced learning. In addition, the inclusion of excessively *high negative rewards can distort the learning curve* of the RL agent. Therefore, the design of the reward function should ensure that the obtained values at each time step are not excessively large to avoid introducing training bias.

Conclusion

This study highlights the application and importance of RL optimization in efficiently managing complex DES. By comparing with conventional boiler firing, feedwater controls, and thermal load-following CHP control, we assessed the potential savings achievable through an RL-based optimizer. RL-based control exhibits 15.99% increased electricity generation along with 35% increased fuel cost, yet the net energy cost savings achievable is 32.30% in comparison to the baseline case. While in its early stages, this investigation shows promise in enhancing the performance of the RL agent through additional training and refinement of reward functions. Moreover, this work had limited demonstration of demand response actions, while the integration of thermal and electric storage and renewable generation could highlight the RL's capabilities of handling multi-state multi-action control problems. Furthermore, the Modelica models developed in this research are currently proposed for open-source release in the MBL. This release will facilitate the adoption of model-based design and optimization, as well as act as virtual environments for model-free optimization strategies for grid-integrated DES.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Advanced Manufacturing Office, Award Number DE-EE0009139. The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government. The authors acknowledge Prof. James D. Freihaut and the Solar Turbines Engine Company for providing the performance curve generation tool to support model implementation.

References

Åkesson, Johan, K-E Årzén, Magnus Gäfvert, Tove Bergdahl, and Hubertus Tummescheit. 2010. "Modeling and optimization with Optimica and JModelica. org—Languages and tools for solving large-scale dynamic optimization problems." *Computers & Chemical Engineering* 34 (11): 1737–1749.

Allegrini, Jonas, Kristina Orehounig, Georgios Mavromatidis, Florian Ruesch, Viktor Dorer, and Ralph Evins. 2015. "A review of modelling approaches and tools for the simulation of district-scale energy systems." *Renewable and Sustainable Energy Reviews* 52:1391–1404.

Bataille, Chris, Henri Waisman, Michel Colombier, Laura Segafredo, and Jim Williams. 2016. The deep decarbonization pathways project (DDPP): insights and emerging issues.

Blockwitz, Torsten, Martin Otter, Johan Akesson, Martin Arnold, Christoph Clauss, Hilding Elmqvist, Markus Friedrich, Andreas Junghanns, Jakob Mauss, Dietmar Neumerkel, et al. 2012. "Functional mockup interface 2.0: The standard for tool independent exchange of simulation models." *Proceedings*.

Blum, David, Javier Arroyo, Sen Huang, Ján Drgoňa, Filip Jorissen, Harald Taxt Walnum, Yan Chen, Kyle Benne, Draguna Vrabie, Michael Wetter, et al. 2021. "Building optimization testing framework (BOPTEST) for simulation-based benchmarking of control strategies in buildings." *Journal of Building Performance Simulation* 14 (5): 586–610.

Ceusters, Glenn, Román Cantú Rodríguez, Alberte Bouso García, Rüdiger Franke, Geert Deconinck, Lieve Helsen, Ann Nowé, Maarten Messagie, and Luis Ramirez Camargo. 2021. "Model-predictive control and reinforcement learning in multi-energy system case studies." *Applied Energy* 303:117634.

Cohen, Scott D, Alan C Hindmarsh, and Paul F Dubois. 1996. "CVODE, a stiff/nonstiff ODE solver in C." *Computers in physics* 10 (2): 138–143.

Heendeniya, Charitha Buddhika, Andreas Sumper, and Ursula Eicker. 2020. "The multi-energy system co-planning of nearly zero-energy districts–Status-quo and future research potential." *Applied energy* 267:114953.

Hermansen, Rune, Kevin Smith, Jan Eric Thorsen, Jiawei Wang, and Yi Zong. 2022. "Model predictive control for a heat booster substation in ultra low temperature district heating systems." *Energy* 238:121631.

Hinkelman, Kathryn, Saranya Anbarasu, Michael Wetter, Antoine Gautier, Baptiste Ravache, and Wangda Zuo. 2022a. "Towards Open-Source Modelica Models for Steam-Based District Heating Systems." *2022 Open Source Modelling and Simulation of Energy Systems (OSMSES)*. IEEE, 1–6.

Hinkelman, Kathryn, Saranya Anbarasu, Michael Wetter, Antoine Gautier, and Wangda Zuo. 2022b. "A fast and accurate modeling approach for water and steam thermodynamics with practical applications in district heating system simulation." *Energy* 254:124227.

Hinkelman, Kathryn, Jing Wang, Wangda Zuo, Antoine Gautier, Michael Wetter, Chengliang Fan, and Nicholas Long. 2022c. "Modelica-based modeling and simulation of district cooling systems: A case study." *Applied Energy* 311:118654.

Klemm, Christian, and Peter Vennemann. 2021. "Modeling and optimization of multi-energy systems in mixed-use districts: A review of existing methods and approaches." *Renewable and Sustainable Energy Reviews* 135:110206.

Lee, Donghun, Seok Mann Yoon, Jaeseung Lee, Kwanho Kim, and Sang Hwa Song. 2020. "Applying Deep Learning to the Heat Production Planning Problem in a District Heating System." *Energies* 13 (24): 6641.

Lund, Rasmus, Danica Djuric Ilic, and Louise Trygg. 2016. "Socioeconomic potential for introducing large-scale heat pumps in district heating in Denmark." *Journal of Cleaner Production* 139:219–229.

Müller, Dirk, Moritz Lauster, Ana Constantin, Marcus Fuchs, and Peter Remmen. 2016. "AixLib-An open-source modelica library within the IEA-EBC annex 60 framework." *BauSIM 2016*, pp. 3–9.

Raffin, Antonin, Ashley Hill, Maximilian Ernestus, Adam Gleave, Anssi Kanervisto, and Noah Dormann. 2019. Stable baselines3.

Revesz, Akos, Phil Jones, Chris Dunham, Gareth Davies, Catarina Marques, Rodrigo Matabuena, Jim Scott, and Graeme Maidment. 2020. "Developing novel 5th generation district energy networks." *Energy* 201:117389.

Saloux, Etienne, and José A Candanedo. 2020. "Optimal rule-based control for the management of thermal energy storage in a Canadian solar district heating system." *Solar Energy* 207:1191–1201.

Schulman, John, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. "Proximal policy optimization algorithms." *arXiv preprint arXiv:1707.06347*.

Sutton, Richard S, and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

Testasecca, Tancredi, Marilena Lazzaro, Elissaios Sarmas, and Stathis Stamatopoulos. 2023. "Recent advances on data-driven services for smart energy systems optimization and pro-active management." *2023 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv)*. IEEE, 146–151.

Verrilli, Francesca, Seshadhri Srinivasan, Giovanni Gambino, Michele Canelli, Mikko Himanka, Carmen Del Vecchio, Maurizio Sasso, and Luigi Glielmo. 2016. "Model predictive control-based optimal operations of district heating system with thermal energy storage and flexible loads." *IEEE Transactions on Automation Science and Engineering* 14 (2): 547–557.

Wang, Dan, Cheng Gao, Yuying Sun, Wei Wang, and Shihao Zhu. 2023. "Reinforcement learning control strategy for differential pressure setpoint in large-scale multi-source looped district cooling system." *Energy and Buildings* 282:112778.

Wetter, Michael, and Christoph van Treeck. 2017. "IEA EBC Annex 60: new generation computing tools for building and community energy systems." *The Regents of the University of California and RWTH Aachen University*.

Wetter, Michael, Wangda Zuo, Thierry S Nouidui, and Xiufeng Pang. 2014. "Modelica buildings library." *Journal of Building Performance Simulation* 7 (4): 253–270.