SimBuild
2024

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

# Applications in CityLearn Gym Environment for Multi-Objective Control Benchmarking in Grid-Interactive Buildings and Districts

Kingsley Nweye[1] and Zoltan Nagy[1]

[1]The University of Texas at Austin, Austin, Texas, USA

## Abstract

It is challenging to coordinate multiple distributed energy resources in a single or multiple buildings to ensure efficient and flexible operation. Advanced control algorithms such as model predictive control and reinforcement learning control provide solutions to this problem by effectively managing a distribution of distributed energy resource control tasks while adapting to unique building characteristics, and cooperating towards improving multi-objective key performance indicator. Yet, a research gap for advanced control adoption is the ability to benchmark algorithm performance. CityLearn addresses this gap an open-source Gym environment for the easy implementation and benchmarking of simple rule-based control and advanced algorithms that has an advantage of modeling simplicity, multi-agent control, district-level objectives, and control resiliency assessment. Here we demonstrate the functionalities of CityLearn using 17 different building control problems that have varying complexity with respect to the number of controllable distributed energy resources in buildings, the simplicity of the control algorithm, the control objective, and district size.

## Introduction

The electricity grid in the United States (U.S.) is undergoing system-wide changes due to the electrification of buildings and transport systems with the goal of creating a robust and resilient electricity network and reducing carbon footprint (U.S. Department of Energy 2021). Buildings are increasingly seen as participatory actors within the electricity market with the addition of on-site renewable energy sources (RESs) and energy storage systems (ESSs) (Neukomm, Nubbe, and Fares 2019). However, these distributed energy resources (DERs) cause a deviation from the business-as-usual energy profile, introduce new peaks, and increase the complexity of building energy demand prediction. At the urban scale, this dynamic and increasing building energy demand translates to periods of critical demand peak that threaten grid resiliency (Do et al. 2023). Furthermore, extreme-weather events such as heat waves and winter storms due to climate change, exacerbate the risk of power outages (Wei et al. 2023).

DERs including RESs, ESSs, and heat pumps that are installed at the point of consumption i.e., buildings, can provide the grid with energy flexibility in response to grid signals, changing weather conditions, or occupant preferences (Jensen et al. 2017). Temporary load shedding, load shifting to off-peak hours, load modulation, and management of renewable power generation are several ways in which buildings can activate their flexibility (Neukomm, Nubbe, and Fares 2019). It is, however, challenging to coordinate multiple DERs in a single or multiple buildings to ensure efficient and flexible operation. Advanced control algorithms such as model predictive control (MPC) (Drgoňa et al. 2020) and reinforcement learning control (RLC) (Nagy et al. 2023) provide solutions to this problem by effectively carrying out a distribution of DER control tasks while adapting to unique building characteristics, and cooperating towards improving multi-objective key performance indicators (KPIs).

A research gap in the adoption of advanced control in demand response (DR) is the ability to benchmark algorithm performance in buildings (Vázquez-Canteli and Nagy 2019). Physics-based models accurately capture the thermodynamics in buildings and are the industry standard for assessing the impact of control in the built environment. However, they require domain knowledge and attention to detail to model building energy systems (BESs) as well as significant effort to integrate standardized advanced control libraries in order to co-simulate and benchmark control algorithms. Building emulators provide various levels of abstractions of BESs, which enables the designer to focus on the control implementation. Such emulators include Energym (Scharnhorst et al. 2021), and BOPTEST (Blum et al. 2021), which provides high-fidelity energy models for control algorithm benchmarking. Although these emulators take advantage of robust simulation engines such as EnergyPlus and Modelica, their dependency on such engines raises the entry level for users. Moreover, they are designed for specific system-level or building-level environments and do not allow for district or neighborhood-level control or objectives. The more recent DOPTEST (Arroyo et al. 2023) framework builds upon BOPTEST to support district-level and multi-agent control but requires compute-intensive co-simulation and is limited to building and district environments provided within the framework.

CityLearn is an open-source Gym environment for the easy implementation and benchmarking of rule-based control (RBC), RLC and MPC algorithms in a demand response setting (Nweye et al. 2023a). CityLearn is used to reshape the aggregated curve of electricity demand by controlling the energy storage of a diverse set of buildings in a district, thus allows for multi-agent control and district-level objectives. The initial release of CityLearn was designed to only manage ESS control for load shifting without altering indoor temperature conditions by making use of pre-defined building thermal loads (Vázquez-Canteli et al. 2019). Thus, the environment excluded a temperature dynamics model since ideal loads were maintained. This approach, however, left out the flexibility potential of electric-source heating ventilation and air conditioning (HVAC) systems e.g., heat pumps to provide load shedding energy flexibility, preheating, and pre-cooling to improve energy efficiency and maintain comfort. The latest CityLearn v2 release, at the time of writing, incorporates a temperature dynamics model that allows for heat pump power control to provide partial load satisfaction.

Thus, our work here addresses the aforementioned research gap of benchmarking algorithm performance in buildings by providing environment and control examples as well as their implementation source code for members of the building control community whom are interested in using the latest CityLearn version to benchmark control algorithms for BES management. We demonstrate different control tasks of differing complexity that can be tackled in CityLearn v2 where complexity refers to (1) the number of controllable DERs present in a building including domestic hot water (DHW) thermal energy storage (TES), battery energy storage system (BESS) when paired with photovoltaic (PV) system for self-generation and heat pump, (2) the simplicity of the control algorithm i.e., explainable RBC or adaptive but black-box RLC, (3) the control objective including reduction in electricity consumption, cost, greenhouse gas (GHG) emission, discomfort, peak demand, and their combinations, and (4) the size of the district i.e., number of buildings. While the provided examples make use of RBC and RLC algorithms, these can be substituted with other control algorithms such as MPC. The remainder of our paper is structured as follows: we provide details about CityLearn's functionalities and how it interfaces with control agents in the CityLearn section and describe our control simulation examples as well as the dataset used in these examples in the Application section. The results of the simulations are presented in the Results section and discussed in the Discussion section. Finally, we summarize our findings and set the stage for future work in the Conclusion section.

## CityLearn

The CityLearn environment (Figure 1) consists of simplified building energy models that contain HVAC systems (heat pumps and electric heaters) and optional BESS and TES. The dynamics of these energy systems are based on reduced-order thermodynamic models that are documented in (Vázquez-Canteli et al. 2019). Each building's space cooling, space heating and DHW heating loads are independently satisfied through air-to-water heat pumps. Alternatively, heat pumps may be replaced with electric heaters to satisfy space and DHW heating loads. TESs are charged by the HVAC system that satisfies the end-use, which the stored energy services. All HVAC systems as well as plug loads consume electricity from any of the available electricity sources including the grid, PV system, and BESS. RBCs, RLC or MPC agent(s) are then used to manage load shifting in the buildings by determining how much energy to store or release at each control time step. The control architecture is either one agent to many buildings (centralized) or one agent to one building (independent) with optional information sharing amongst agents to achieve cooperative or competitive objectives.

The agents can also, control the available power from the HVAC system to shed space thermal loads, preheat, or pre-cool the buildings. The consequence of the controlled power on the indoor dry-bulb temperature, i.e., the building dynamics, is modeled using a long short-term memory (LSTM) surrogate model based on the methodology by Pinto et al. (Pinto, Deltetto, and Capozzoli 2021). This methodology leverages EnergyPlus to generate training data including cooling or heating load and indoor temperature observations by varying the proportion of cooling or heating ideal load satisfied in conditioned zones in a building. These observations are combined with temporal and weather variables for use in

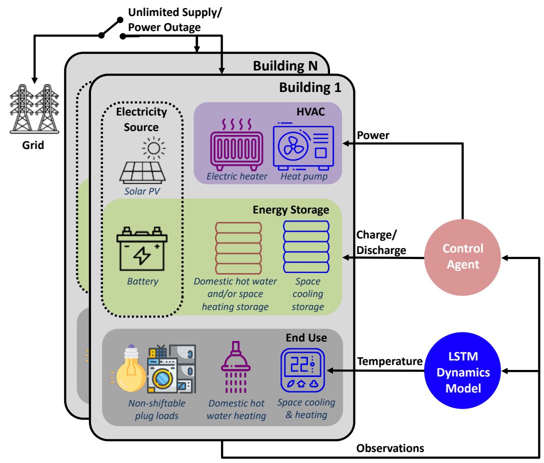



Figure 1: CityLearn environment and control interaction.

training the LSTM model. The trained model is loaded during environment initialization and is used to predict a building's indoor dry-bulb temperature given a sequence of previously observed temperatures, loads, weather and temporal variables. With the provision of OpenStudio models in the End-Use Load Profiles (EULP) dataset (Wilson et al. 2022) that are convertible to EnergyPlus models as well as the work in (Nweye et al. 2023b) to apply the EULP dataset to create CityLearn input datasets from it, temperature dynamics models can be developed for a plethora of building configurations and locations across the U.S. for use in control benchmarking in CityLearn.

CityLearn also provides the functionality to assess the resiliency of control algorithms during power outage events. During a power outage event, the grid is unable to provide buildings with electricity and agents can only make use of the flexibility provided by DERs in buildings to satisfy loads otherwise, risk thermal discomfort and unserved energy during the event. Whereas, during normal operation, there is unlimited supply from the grid. The outage signals are either statically defined as a time series or generated using some stochastic model. We provide a stochastic power outage model based on System Average Interruption Frequency Index (SAIFI) and Customer Average Interruption Duration Index (CAIDI) distribution system reliability metrics (IEEE 2012) such that it is customizable and adaptable by providing location-specific metrics. The stochastic signals are generated by sampling $n$ instances from a binomial distribution to select days that experience power outage where the probability of a day having an outage, $p$, is the ratio of SAIFI to number of days in a year (365). The start time step index for the outage on each day is then randomly selected from a uniform distribution. Finally, the duration of each power outage event is set by sampling from an exponential distribution with a scale set to CAIDI.

Being an open-source environment, CityLearn can be customized to consider models of other DERs beyond what is contained in the as-provided code base including, electric vehicles (EVs), solar collector, and other controllable plug loads. It also comes prepackaged with a number of environment configuration datasets that have been used in The CityLearn Challenge (Nweye et al. 2022b) and its code base[1] has been extensively documented to support its application.

## Application

In the following subsections, we describe the dataset used to generate the environment for each control simulation example as well as descriptions of the different environment, objective, and control algorithm combinations that make up the simulations. The source code for our work is made available on GitHub[2].

### Dataset

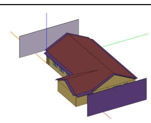
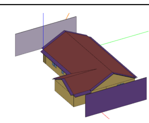
We use a dataset of two synthetic residential single-family buildings, B1 and B2, in Austin, Texas, U.S. that are summarized in Table 1. These buildings are sampled from building energy models in (Wilson et al. 2022) that are representative of the residential and commercial building in the U.S.. Both buildings have the same geometry and floor area but different load profile, baseline monthly consumption, construction decade, and BES sizing. Each building has a DHW TES and BESS-PV system for load shifting as well as controllable heat pump for load shedding. The environment is based on a one-month summer period of June 2018 thus, the heat pump is used only for cooling load satisfaction. An electric heater is used to satisfy DHW heating load and charge the DHW TES. The DHW TES has equal maximum charge and discharge rates equivalent to the electric heater's nominal power. The DHW TES capacity is conservatively sized to equal half the maximum hourly DHW heating demand in the one-month period, whereas the heat pump and electric heater are sized to satisfy the maximum hourly cooling and DHW heating loads, respectively. The BESS is the equivalent of a real-world manufacturer's specifications with equal max-

[1] https://www.citylearn.net

[2] https://github.com/intelligent-environments-lab/simbuild-2024-citylearn/tree/v0.0.1

imum charge and discharge nominal power of 3.3kW for B1 and 1.6kW for B2, and a 20% depth-of-discharge. The PV system is conservatively sized for a 21.5% zero-net energy (ZNE) scenario.

Table 1: Environment building metadata. The average daily profile in the train period for cases with PV generation (solid line) and without PV generation (dotted line) are shown as well as the one-month electricity consumption without control and PV generation.

| | **B1** | **B2** |
|---|---|---|
| Geometry |  |  |
| Avg. daily profile |  |  |
| Consumption (kWh) | 791.3 | 1293.8 |
| Vintage | 1980s | 1970s |
| Floor area ($m^2$) | 157.0 | 157.0 |
| Heat pump (kW) | 2.3 | 2.8 |
| DHW heater (kW) | 3.7 | 6.3 |
| DHW TES (kWh) | 1.7 | 2.8 |
| BESS (kWh) | 4.0 | 3.3 |
| BESS (kW) | 3.3 | 1.6 |
| PV (kW) | 1.2 | 2.4 |

Other supplementary datasets are 2018 weather data from the Austin Camp Mabry weather station that is provided by (Wilson et al. 2022), Austin Energy time-of-use (ToU) rates in Table 2 that is reflective of its 2018 ToU pilot program (City of Austin 2017) and excludes any other static base rates. We also leverage the Electric Reliability Council of Texas (ERCOT) fuel mix dataset (Electric Reliability Council of Texas 2021) and reported GHG emissions for different energy sources in (Moomaw et al. 2011) to calculate hourly grid $CO_2$e intensity (kg$CO_2$e/kWh).

Using the framework in (Nweye et al. 2023b), we convert the building and supplementary datasets to a CityLearn input dataset. It takes five minutes on average for us to train a single building's LSTM temperature dynamics model on a 16GB RAM, Intel Core i7 1255U CPU machine. We use the initial 13 days in the one-month period for control agent hyperparameter tuning and training (train period) while the remaining 17 days are used for inference and evaluation of the control objectives (test period).

Table 2: ToU electricity rate.

| | **Rate ($/kWh)** | |
|---|---|---|
| **Time** | **Weekday** | **Weekend** |
| 7 AM - 3 PM (Mid-Peak) | 0.0291 | 0.0289 |
| 3 PM - 6 PM (On-Peak) | 0.0587 | 0.0289 |
| 6 PM - 10 PM (Mid-Peak) | 0.0291 | 0.0289 |
| 10 PM - 7 AM (Off-Peak) | 0.0289 | 0.0289 |

## Control Simulations

We demonstrate the use of either RBC or RLC to manage the ESSs and heat pump in the buildings. The BESS and DHW TES control actions received from the agent are the proportions of their capacities to be charged or discharged $\in [-1, 1]$ whereas, that of the heat pump is the proportion of its nominal power $\in [0, 1]$.

The simulations are designed to minimize at least, one of four building-level objectives and one district-level objective (Table 3), including cost (Equation (1)), emissions (Equation (2)), discomfort (Equation (3)), consumption (Equation (4)), and average daily peak (Equation (5)). In the aforementioned equations, $t$ is the simulation time step, $n$ is the total number of time steps, $d$ is the day-of-year index, $h$ is the number of time steps in a day, which in our case is $h = 24$. $e$ is the building-level electricity consumption in kWh, $P$ is district-level electricity demand in kW, $R$ is the electricity rate in $/kWh, $G$ is the $CO_2$e intensity in kg$CO_2$e/kWh, $T$ is the indoor dry-bulb temperature in °C, and $T_{\mathrm{spt}}$ is the indoor dry-bulb temperature setpoint.

There is a total of 17 environment, objective, and control algorithm combinations we evaluate that vary in the type of algorithm complexity, reward function for RLC algorithms, control objective, and controlled DERs including DHW TES, BESS-PV system, and heat pump as summarized in Table 4.

## Baseline (No Control)

There are two baseline configurations, *x-b1_b2-x-x* and *x-b1_b2-x-pv* that include both B1 and B2 in the environment where there are no controlled DERs. However, *x-b1_b2-x-pv* includes a PV system in each building to augment electricity provision from the grid.

Table 3: Control objectives to minimize

| Objective | Definition | |
|---|---|---|
| Cost | $\sum_{t=0}^{n-1} \max(e(t),0) \times R(t)$ | (1) |
| Emissions | $\sum_{t=0}^{n-1} \max(e(t),0) \times G(t)$ | (2) |
| Discomfort | $\sum_{t=0}^{n-1} \lvert T(t) - T_{\text{spt}}(t) \rvert$ | (3) |
| Consumption | $\sum_{t=0}^{n-1} \max(e(t),0)$ | (4) |
| Avg. Daily Peak | $\frac{\left(\sum_{d=0}^{(n \div h)-1} \max(P(d))\right) \cdot h}{n}$ | (5) |

### Rule-Based Control

We define seven RBC-based configurations. *rbc-b1-c-dhw* and *rbc-b1-e-dhw* include only building B1 equipped with DHW TES control but differ in their control objectives, with the former being cost reduction and the latter being emissions reduction. Accordingly, the RBC agents are tailored to meet these objectives, where we tune a cost-reduction-based RBC according to the rates in Table 2. On weekdays, the RBC is set to charge the ESS to capacity during the off-peak hours of 10 PM to 7 AM, discharge 50% of its capacity during the on-peak period of 3 PM to 6 PM and discharge the remaining energy at other times. During the weekend, when a cheap flat rate is used, the ESS is charged at a constant rate.

We tune the emission-reduction-based RBC by qualitatively analyzing the ERCOT average daily $CO_2$e intensity profile during the train period. The grid $CO_2$e intensity is the least prior to 8 AM, peaks around 1 PM but sustains a high intensity between noon and 11 PM. Thus, the emission-reduction-based RBC is set to fully charge the ESS at a constant rate prior to 8 AM and discharge the stored energy at a constant rate between 12 PM and 11 PM.

The *rbc-b1-c-bess_pv* and *rbc-b1-e-bess_pv* configurations replace the DHW TES with a BESS-PV system whereas, the *rbc-b1-c-dhw_bess_pv*, and *rbc-b1-e-dhw_bess_pv* augment the DHW TES with the BESS-PV system. The *rbc-b1_b2-p-bess_pv* configuration uses both buildings B1 and B2 with a district-level objective of average daily peak reduction (Equation (5)) and makes use of a peak-reduction-based RBC to control the BESS. We tune the RBC by inferring the peak hours in each building's daily profile shown in Table 1 such that it charges the ESS to capacity before 6 AM at steady and then, discharges it at a steady rate between 6 AM and 11 PM.

### Reinforcement Learning Control

We design eight RLC configurations, of which seven are the RLC equivalent of the seven RLC configurations. We make use of the soft actor-critic (SAC) RLC algorithm (Haarnoja et al. 2018) and take advantage of its default implementation in the Stable-Baselines3 Python package (Raffin et al. 2021). Each configuration environment is used to train an agent on 150 epochs of the 13-day train period.

We define a reward function in Table 5 for each RLC objective in Table 4. The *rlc-b1-c-dhw*, *rlc-b1-c-bess_pv*, and *rlc-b1-c-dhw_bess_pv* configurations have the same cost objective and reward function (Equation (6)). Equation (6) is similar to the objective function defined in, Equation (1) with the exception that it only considers the cost at the current time step. Given that the goal in training an RLC agent is to maximize the cumulative reward, we apply a negative sign to the calculated reward so that large costs have lower rewards. A similar approach is taken for the reward function used in the *rlc-b1-e-dhw*, *rlc-b1-e-bess_pv*, and *rlc-b1-e-dhw_bess_pv* configurations where the emission reward function (Equation (7)) is the negative equivalent of Equation (2) for a single time step.

The *rlc-b1-d_o-hp* configuration is multi-objective where, the goal is to minimize electricity consumption (Equation (4)) while maintaining comfort (Equation (3)). We encode both objectives in the discomfort and consumption reward function defined in Equation (8) where the penalized delta between $T$ and $T_{\text{spt}}(t)$ is scaled up by a factor of $m$, when $T < T_{\text{spt}}(t)$ to prevent the RLC agent from over-cooling the building which has an adverse effect of increased electricity consumption.

The *rlc-b1_b2-p-bess_pv* configuration only differs from *rbc-b1_b2-p-bess_pv* in control algorithm. *rlc-b1_b2-p-bess_pv* utilizes the SAC-glsrlc agent and average daily peak reward function (Equation (9)) to minimize average daily peak (Equation (5)) where the reward function is set to penalize high electricity demand which by proxy penalizes peaks.

We set each RLC-based configuration to use a subset of

Table 4: Environment, objective, and control algorithm combinations that vary in the type of control algorithm, reward function for RLC algorithms, control objective, and controlled DERs.

| Configuration ID | Bldgs. | Objective | Reward | Control Algo. | DHW | BESS | PV | Heat Pump |
|---|---|---|---|---|---|---|---|---|
| *x-b1_b2-x-x* | 2 | No Control | - | - | | | | |
| *x-b1_b2-x-pv* | 2 | No Control | - | - | | | ✓ | |
| *rbc-b1-c-dhw* | 1 | Cost | - | Cost-RBC | ✓ | | | |
| *rbc-b1-e-dhw* | 1 | Emissions | - | Emission-RBC | ✓ | | | |
| *rbc-b1-c-bess_pv* | 1 | Cost | - | Cost-RBC | | ✓ | ✓ | |
| *rbc-b1-e-bess_pv* | 1 | Emissions | - | Emission-RBC | | ✓ | ✓ | |
| *rbc-b1-c-dhw_bess_pv* | 1 | Cost | - | Cost-RBC | ✓ | ✓ | ✓ | |
| *rbc-b1-e-dhw_bess_pv* | 1 | Emissions | - | Emission-RBC | ✓ | ✓ | ✓ | |
| *rbc-b1_b2-p-bess_pv* | 2 | Peak | - | Peak-RBC | | ✓ | ✓ | |
| *rlc-b1-c-dhw* | 1 | Cost | Equation (6) | SAC-RLC | ✓ | | | |
| *rlc-b1-e-dhw* | 1 | Emissions | Equation (7) | SAC-RLC | ✓ | | | |
| *rlc-b1-c-bess_pv* | 1 | Cost | Equation (6) | SAC-RLC | | ✓ | ✓ | |
| *rlc-b1-e-bess_pv* | 1 | Emissions | Equation (7) | SAC-RLC | | ✓ | ✓ | |
| *rlc-b1-c-dhw_bess_pv* | 1 | Cost | Equation (6) | SAC-RLC | ✓ | ✓ | ✓ | |
| *rlc-b1-e-dhw_bess_pv* | 1 | Emissions | Equation (7) | SAC-RLC | ✓ | ✓ | ✓ | |
| *rlc-b1-d_o-hp* | 1 | Discomfort & Consumption | Equation (8) | SAC-RLC | | | | ✓ |
| *rlc-b1_b2-p-bess_pv* | 2 | Peak | Equation (9) | SAC-RLC | | ✓ | ✓ | |

Table 5: Reward functions.

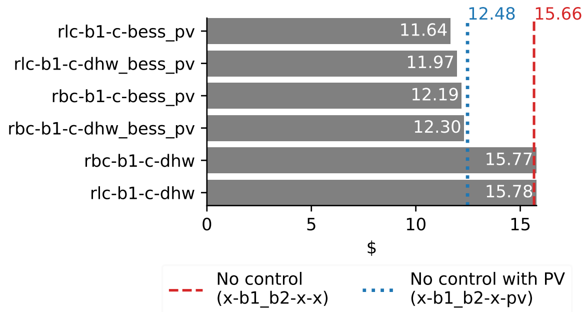
| Penalty | Definition | |
|---|---|---|
| Cost | $-\max(e(t),0) \times R(t)$ | (6) |
| Emissions | $-\max(e(t),0) \times G(t)$ | (7) |
| Discomfort & Cons. | $\begin{cases} -m \times \lvert T(t) - T_{\text{spt}}(t) \rvert, \text{ if } T(t) < T_{\text{spt}}(t) \\ -\lvert T(t) - T_{\text{spt}}(t) \rvert, \text{ otherwise} \end{cases}$ | (8) |
| Avg. Daily Peak | $-\max(P(t),0)$ | (9) |

12 observations. The hour and day-of-week observations are used in all simulations. Net electricity consumption is excluded in the configurations that include heat pump control and discomfort objective. The electricity rate and $CO_2$e intensity observations are only relevant to configurations that seek to minimize cost and emissions, respectively. Solar generation and BESS state-of-charge (SoC) observations are included in environments that have BESS-PV systems in the buildings whereas DHW TES SoC is observed only in environments that have DHW TES. Outdoor and indoor dry-bulb temperature, indoor dry-bulb temperature setpoint and the absolute delta between indoor dry-bulb temperature and indoor dry-bulb temperature setpoint are active only in environments with heat pump control.

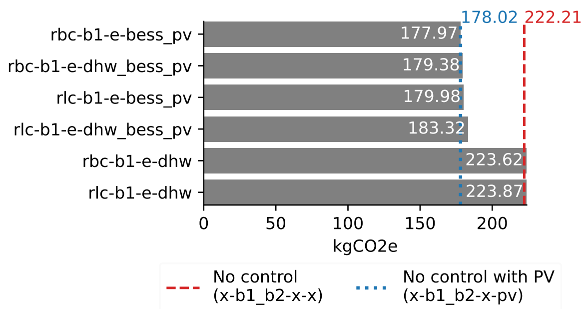
## Results

### Single Building Cost or Emissions Objective

Figure 2 shows cost ($) and emissions (kg$CO_2$e) in building B1 from either RBC or RLC of DHW TES, BESS-PV system or both when the control objective is cost (Figure 2a) or emission (Figure 2b), compared to baseline configurations with no control (*x-b1_b2-x-x*) and no control but solar generation (*x-b1_b2-x-pv*) in the 17-day test period. We show that the PV system advantage in terms of cost and emissions reduction is approximately 20.0%, similar to the 21.5% ZNE that the PV system was sized for. There is no cost or emissions reduction benefit from DHW TES without PV system control when either an RBC or RLC agent is used. The largest cost savings compared to *x-b1_b2-x-pv* is $0.83 (6.7%) that is provided by the reinforcement learning (RL)-controlled BESS-PV system (*rlc-b1-c-bess_pv*). Including a DHW-TES to the BESS-PV environment (*rlc-b1-c-dhw_bess_pv*) slightly increases the cost but still performs better than the rule-base-controlled BESS-PV system. Both RBC configurations that include solar generation in the building have higher costs than their RLC alternative but still perform better than the baseline in magnitudes of 2.3% and 1.4% for *rbc-b1-c-bess_pv* and *rbc-b1-c-dhw_bess_pv*.




(a) Cost ($).




(b) Emissions (kg$CO_2$e).

Figure 2: Cost ($) and emissions (kg$CO_2$e) from either RBC or RLC of DHW TES, BESS-PV system or both when the control objective is cost or emission. The dashed red line shows the cost for a no-control scenario and the dotted blue line shows the cost for a no-control scenario but with solar generation to augment electricity from the grid.

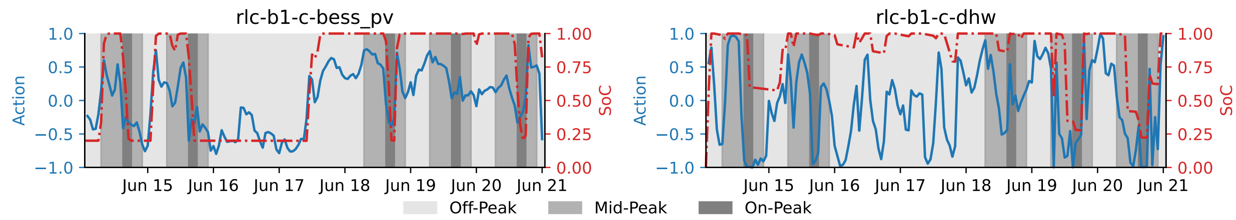
In contrast and shown in Figure 2b, neither RBC nor RLC provides emissions reduction irrespective of the available ESSs when paired with a PV system and our manually tuned RBC performs marginally better than the advanced RLC agent for an emissions reduction objective. The *rlc-b1-e-dhw_bess_pv* environment in fact, increases emissions by approximately 3.0%.

Figure 3 shows the RLC actions in the initial seven days of the test period for the configurations that have one ESS with cost reduction objective (*rlc-b1-c-dhw* and *rlc-b1-c-dhw_bess_pv*), or emissions reduction objective (*rlc-b1-e-dhw* and *rlc-b1-e-dhw_bess_pv*). The resulting changes in the ESS SoC as a consequence of these actions are also shown as well as shaded regions to indicate the ToU as defined in Table 2 as well as the $CO_2$e intensity. The SoC trends in Figure 3a show that with the cost reward function (Equation (6)), the RLC agent learns to charge the ESSs during the off-peak period while discharging mainly during the on-peak period. However, the charge and discharge rate are rapid, reaching full charge and discharge depth within few time steps. Although, the agent calls for charging during the mid-peak period, this ToU coincides with when the ESS is charged to capacity i.e., SoC $= 1$ thus, has no effect on the SoC. There is also an unusual call to discharge throughout June 16, a Saturday, when electricity is cheapest. Compared to the BESS, the DHW TES rarely discharges its full capacity as the agent calls for charge action, even when the SoC is well above 75.0%. There is also no change in DHW TES SoC in some instances where the agent calls for energy discharge.

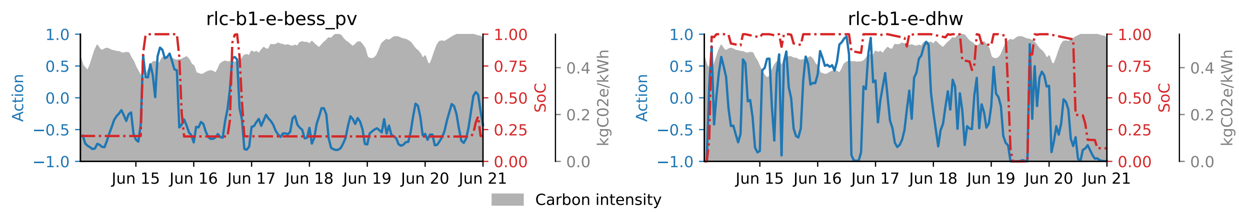
The results for an emissions reduction objective, depicted in Figure 3b, show persistent call for discharge of the BESS by the RLC agent for almost four consecutive days even though the BESS has reached its depth-of-discharge. The agent does take advantage of the lower $CO_2$e intensity to charge however, there is a very mild variation in carbon intensity for the evaluated period as there are no distinct periods of significantly low emissions to take advantage of. In contrast, the RLC agent controlling the DHW TES shows no clear pattern in selected actions as it charges and discharges at random. We also observe a similar SoC trend as with the cost objective where the TES rarely depletes below 75.0% SoC.

**Single Building Discomfort and Consumption Objective**

In Table 6, we show the effect of varying the multiplier, $m$ in the discomfort and consumption reward function on discomfort, average discomfort when over-cooling ($\overline{T_\Delta} < 0$) and under-cooling ($\overline{T_\Delta} > 0$), and consumption objectives when used in heat pump control environment, *rlc-b1-d_o-hp*. The percent changes in objective by setting $m = 3$, $m = 6$, and $m = 12$ are compared to the baseline case, $m = 1$, where there is no penalty for over-cooling i.e., objective is reduced to discomfort reduction alone. We see that by setting $m = 3$ to activate the multi-objective reward, we reduce consumption by 3.0% but also discomfort by 3.0% through 50.0% decrease in over-cooling. In contrast, $m = 3$ increases under-cooling by only 20.0%. Larger increments in $m$ to $m = 6$ and $m = 12$ show an increase in discomfort of almost 100.0% compared to the baseline but up to 12.0% decrease in consumption. However, the average temperature difference between the indoor dry-bulb temperature and setpoint for the selected $m$ values, irrespective of over-cooling or under-cooling, is less than 1.5C.




(a) BESS-PV system (right) and DHW TES (left) control with cost objective.




(b) BESS-PV system (right) and DHW TES (left) control with emissions objective.

Figure 3: RLC action and consequent ESS SoC trend in the initial seven days of the two-week evaluation period when the control objective is to either minimize cost or emissions.

Table 6: Effect of varying $m$ in discomfort and consumption reward function, Equation (9), on discomfort, average discomfort when over-cooling ($\overline{T_\Delta} < 0$) and under-cooling ($\overline{T_\Delta} > 0$), and consumption objectives when used in heat pump control environment, *rlc-b1-d_o-hp*. The percent change in objective value, by varying $m$ is compared to the baseline case, $m = 1$, where there is no penalty for increased consumption from over-cooling. Improvement in an objective is highlighted in blue while deterioration is highlighted in red.

| $m$ | Discomfort (°Ch) | Over-cool ($\overline{T_\Delta} < 0$) (°C) | Under-cool ($\overline{T_\Delta} > 0$) (°C) | Consumption (kWh) |
|---|---|---|---|---|
| 1 | 201.8 | $0.4 \pm 0.5$ | $0.5 \pm 0.5$ | 384.8 |
| 3 | 195.7 (-3.0%) | $0.2 \pm 0.2$ (-50.0%) | $0.6 \pm 0.5$ (20.0%) | 373.0 (-3.0%) |
| 6 | 315.9 (56.5%) | $0.3 \pm 0.2$ (-25.0%) | $0.9 \pm 0.5$ (80.0%) | 352.5 (-8.3%) |
| 12 | 399.6 (98.0%) | $0.1 \pm 0.4$ (-75.0%) | $1.0 \pm 0.6$ (100.0%) | 338.1 (-12.0%) |

**District Average Daily Peak Objective**

The daily district-level peak load is shown in Figure 4 for the baseline no-control with PV generation (*x-b1_b2-x-pv*) configuration and the two configurations with peak reduction objective but differing control algorithms: *rbc-b1_b2-p-bess_pv* and *rlc-b1_b2-p-bess_pv*. Compared to the baseline, the RBC is able to reduce the peak on ten out of the 17-day test period at an average of 8.4%. In contrast, the RLC only reduces the peak below the baseline on four days at an average of 5.9%. The RBC never peaks higher than the baseline, in contrast to the RLC that exceeds the baseline thrice. There are eleven days when the RBC is able to perform better than the RLC while the RLC outperforms the RBC on only two days. There are four days when neither control approach is able to reduce peak load compared to the baseline. In summary, the RBC provides a 2.6% advantage in average daily peak reduction over RLC. However, neither agent is able to reduce the highest peak on June 28.

## Discussion

Our results for the *rbc-b1-c-dhw*, *rlc-b1-c-dhw*, *rbc-b1-e-dhw*, and *rls-b1-e-dhw* configurations where we shift DHW loads in B1 using an RBC or RLC agent in order to reduce cost and emissions show slightly higher costs and emissions compared to the baseline (Figure 2). We infer that this observation is as a result of our conservatively-sized DHW TES which is only half of the peak hourly

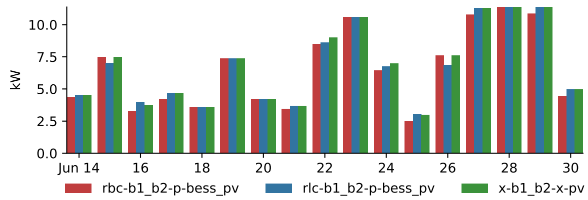



Figure 4: District-level daily peak load for two-building environments where each building has BESS-PV system.

demand. This sizing is a design choice that is characteristic of the dataset and its original curation for use in The CityLearn Challenge 2023[3] as it makes it challenging to leverage largely sized BESs to reduce electricity consumption, costs or emissions. Moreover, the DHW demand in B1 is only 7% of the total building demand, and there is no DHW demand during 330/407 control time steps. Our observation in Figure 3 where the TES almost never depletes below 75.0% capacity irrespective of discharge action by the RLC agent is attributed to the infrequent occurrence of DHW demand and can make association of actions to changes in SoC observations challenging for the agent. In the same vein, the choice or BESS-PV size plays a role in the amount of energy flexibility it is able to provide. In this work, we have used a PV sized for 21.5% ZNE while the BESS capacity is able to meet almost 100.0% of hourly demand. Previous work has however shown that for a single-family grid-interactive building designed for 100.0% ZNE in Fontana, California with similar warm climate as Austin, Texas, up to 27.0% and 37.0% annual cost and emissions reduction respectively achievable (Nweye, Sankaranarayanan, and Nagy 2023) when paired with an RL-controlled BESS.

Although, the ToU rate we use in Table 2 is three-tiered, there is negligible difference in rates value across the tiers, making for a weak cost signal. Likewise, the variance in carbon intensity is less than 1.0% making it also a weak signal. Typically, ToU rates are reflective of how clean the electricity grid is and the time of peak demand. The average daily share of renewable energy sources in the ERCOT grid in June 2018 is estimated at only 20.1% with 5.5% standard deviation (Electric Reliability Council of Texas 2021). This highlights the importance of simultaneously decarbonizing the supply-side as demand-side end uses are electrified or risk an adverse effect of increased emissions. Although no observation forecasts have been used in our work for RLC, providing the agent with cost and $CO_2$e intensity forecast over a short horizon could improve learning and performance.

We observe in Figure 3 that the RLC agents learn to charge when electricity is cheap but take redundant actions, including calling for charging when the ESSs is fully charged or discharging when completely depleted. However, these futile actions have no associated penalty. We consider a reward function that has a penalty for actions that are unable to change the state of the ESS could improve the action quality of the agent.

The work by June et al. on occupant-centric thermostat control (Park and Nagy 2020) has shown reduced setpoint overrides and energy consumption when occupant comfort and preferences are considered in control decisions. It is estimated from field studies that a median of up to 20.0% energy savings are accrued from occupant-centric control (OCC) (Jung and Jazizadeh 2019). CityLearn provides a framework for such OCC by employing any of two methodology. In the first methodology, ideal loads that satisfy comfort are maintained while a control agent learns to improve energy efficiency through ESSs load shifting. The second methodology, as we have shown in this work with heat pump control in the CityLearn v2 environment, is where both comfort and energy efficiency can be learned by way of reward function design. For a cooling season, we show that by penalizing indoor temperature below the setpoint, an agent is able to learn that over-cooling had an adverse effect of increased consumption. This reward function style has been used in the literature by Pinto et al. where exponents on the temperature delta when outside a comfort band are used to teach the agent to prefer under-cooling over over-cooling (Pinto, Deltetto, and Capozzoli 2021).

Our results from average daily peak reduction show better performance from using an RBC agent that has been finely tuned to target energy discharge during peak periods compared to an RLC agent that tries to learn to reduce the peaks using the electricity consumption as a reward signal. In (Nweye et al. 2022a), Nweye et al. show that an expert tuned RBC agent is able to outperform RLC algorithms however, it may be beneficial to take advantage of the expert rules to train the RLC agent offline. This method of training from expert systems been used in simple gaming applications (Argerich, Fürst, and Cheng 2020) and complex HVAC control (Dey et al. 2023).

Our dataset is constructed from the End-Use Load Profiles (EULP) for the U.S. Building Stock dataset (Wilson et al. 2022) where we follow the methodology in (Nweye et al. 2023b) to create our CityLearn environ-

[3] https://www.aicrowd.com/challenges/neurips-2023-citylearn-challenge

ment. The two buildings we use (Table 1) are each reflective of about 242 actual single-family buildings in Austin, Texas and in comparison to 2020 Residential Energy Consumption Survey (RECS) (U.S. Energy Information Administration ), the total floor area of our buildings are near the average residential building floor area of $161\mathrm{m}^2$ in Texas. The estimated average customer electricity consumption reported by Austin Energy for June 2018 is 1,144 kWh (Austin Energy 2023). This estimate is similar to electricity consumption of 1293.8 kWh in building B2 during our June data period however, building B1 has a much lower consumption that is about 30% less than the survey estimate. We attribute this disparity to the way heat pumps are modeled in CityLearn. The CityLearn heat pump model assumes a Carnot cycle with maximum theoretical efficiency that may not accurately reflect the heat pump operation and specification in the original EULP building model. CityLearn is best applied to simple and quick comparative analysis amongst control algorithms at single-building or district levels with minimal design cost. Whereas, an environment that utilizes high-fidelity energy models e.g., BOPTEST (Blum et al. 2021) for single-building control or DOPTEST (Arroyo et al. 2023) for district-control is better suited for analyses that are purposed for scrutiny of absolute energy and cost values. However, the challenge with using these high-fidelity environments is the limited system and building configurations users are restricted to such that implementing buildings from the EULP dataset for large scale building stock studies will incur significant design costs.

Despite its potential, there are still open questions regarding the plug-and-play capabilities, performance, safety of operation, and learning speed of RLC. The CityLearn Challenge is an opportunity to compete in investigating the potential of artificial intelligence (AI) and distributed control systems to tackle multiple problems within the built-environment domain. Other competitions in the energy domain focus on building load predictions (Miller et al. 2020), grid power flow optimization (Holzer et al. 2021) and pathways to building electrification and decarbonization (NYSERDA 2022). While these competitions provided a plethora of solutions that are pertinent to energy supply and demand-side management, The CityLearn Challenge is one of its kind with a focus on BES control and control algorithm benchmarking for demand response (Nweye et al. 2022b). Previous editions of The CityLearn Challenge have investigated transferability of control policies, a realistic implementation of a model-free RL in buildings where training evaluation is done on a single four-year long episode, and BESS control in a real-world ZNE community for energy flexibility maximization. Thus, CityLearn attracts a multidisciplinary participation audience including researchers, industry experts, sustainability enthusiasts and AI hobbyists as an avenue for crowdsourcing solutions to building control problems using RL or any other algorithm of choice.

## Conclusion

Our work here has shown 17 different building control applications in CityLearn. We show that simple rule-based control and complex reinforcement learning control agents managing different distributed energy resource can be benchmarked in a two-building residential district, based on realistic building stock energy models, towards optimizing either singular objective or multi-objective key performance indicators.

For our dataset, the results show that RLC provides further cost reduction compared to RBC while, neither algorithm provides substantial reduction in emissions irrespective of what DERs are available. Also RLC can learn to reduce discomfort, overcooling and consequently consumption when reward function parameters are adequately tuned. Lastly, for our peak reduction objective, RBC provides a 2.6% advantage over RLC.

The applications presented here are by no means an exhaustive demonstration of the CityLearn environment's functionalities, especially its use for control resiliency benchmarking. Instead, we envision that our contribution provides environment and control examples as well as their implementation source code for members of the built environment community whom are interested in using CityLearn to solve their BES control problems.

## References


Argerich, Mauricio Fadel, Jonathan Fürst, and Bin Cheng. 2020. "Tutor4RL: Guiding Reinforcement Learning with External Knowledge."

Arroyo, Javier, Lucas Verleyen, Lucas Bex, Louis Hermans, Muhammad Saeed, Yucun Lu, Joris Depoortere, Attila Bálint, Erik Delarue, Johan Driesen, Geert Deconinck, David Blum, and L. Helsen. 2023, September. *Prototyping the DOPTEST Framework for Simulation-Based Testing of System Integration Strategies in Districts*.

Austin Energy. 2023, April. Residential Average Monthly kWh and Bills.

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

City of Austin. 2017, November. City of Austin Fiscal Year 2018 Electric Tariff.

Dey, Sourav, Thibault Marzullo, Xiangyu Zhang, and Gregor Henze. 2023. “Reinforcement learning building control approach harnessing imitation learning.” *Energy and AI* 14 (October): 100255.

Do, Vivian, Heather McBrien, Nina M. Flores, Alexander J. Northrop, Jeffrey Schlegelmilch, Mathew V. Kiang, and Joan A. Casey. 2023. “Spatiotemporal distribution of power outages with climate events and social vulnerability in the USA.” *Nature Communications* 14 (1): 2470 (April). Number: 1 Publisher: Nature Publishing Group.

Drgoňa, Ján, Javier Arroyo, Iago Cupeiro Figueroa, David Blum, Krzysztof Arendt, Donghun Kim, Enric Perarnau Ollé, Juraj Oravec, Michael Wetter, Draguna L. Vrabie, and Lieve Helsen. 2020. “All you need to know about model predictive control for buildings.” *Annual Reviews in Control* 50 (May): 190–232.

Electric Reliability Council of Texas. 2021, March. Fuel Mix Report: 2007 - 2020.

Haarnoja, Tuomas, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. “Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor.”

Holzer, Jesse, Carleton Coffrin, Christopher DeMarco, Ray Duthu, Stephen Elbert, Scott Greene, Olga Kuchar, Bernard Lesieutre, Hanyue Li, Wai Keung Mak, and others. 2021. “Grid Optimization Competition Challenge 2 Problem Formulation.” Technical Report, Tech. rep. ARPA-E.

IEEE. 2012, May. IEEE Guide for Electric Power Distribution Reliability Indices. Conference Name: IEEE Std 1366-2012 (Revision of IEEE Std 1366-2003).

Jensen, Soren Ostergaard, Anna Marszal-Pomianowska, Roberto Lollini, Wilmer Pasut, Armin Knotzer, Peter Engelmann, Anne Stafford, and Glenn Reynders. 2017. “IEA EBC Annex 67 Energy Flexible Buildings.” *Energy and Buildings* 155:25–34.

Jung, Wooyoung, and Farrokh Jazizadeh. 2019. “Human-in-the-loop HVAC operations: A quantitative review on occupancy, comfort, and energy-efficiency dimensions.” *Applied Energy* 239:1471–1508. Publisher: Elsevier.

Miller, Clayton, Pandarasamy Arjunan, Anjukan Kathirgamanathan, Chun Fu, Jonathan Roth, June Young Park, Chris Balbach, Krishnan Gowri, Zoltan Nagy, Anthony D. Fontanini, and Jeff Haberl. 2020. “The ASHRAE Great Energy Predictor III competition: Overview and results.” *Science and Technology for the Built Environment* 26 (10): 1427–1447 (August). Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/23744731.2020.1795514.

Moomaw, W., P. Burgherr, G. Heath, M. Lenzen, J. Nyboer, and A. Verbruggen. 2011. “Annex II: Methodology.” In *IPCC special report on renewable energy sources and climate change mitigation*, edited by O. Edenhofer, R. Pichs-Madruga, Y. Sokona, K. Seyboth, P. Matschoss, S. Kadner, T. Zwickel, P. Eickemeier, G. Hansen, S. Schlömer, and C. von Stechow. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Nagy, Zoltan, Burak Gunay, Clayton Miller, Jakob Hahn, Mohamed Ouf, Seungjae Lee, Brodie W. Hobson, Tareq Abuimara, Karol Bandurski, Maíra André, Clara-Larissa Lorenz, Sarah Crosby, Bing Dong, Zixin Jiang, Yuzhen Peng, Matteo Favero, June Young Park, Kingsley Nweye, Pedram Nojedehi, Helen Stopps, Lucile Sarran, Connor Brackley, Katherine Bassett, Krissy Govertsen, Nicole Koczorek, and Oliver a Abele. 2023. “Ten questions concerning occupant-centric control and operations.” *Building and Environment*, June, 110518.

Neukomm, Monica, Valerie Nubbe, and Robert Fares. 2019. “Grid-Interactive Efficient Buildings.” April.

Nweye, Kingsley, Kathryn Kaspar, Giacomo Buscemi, Giuseppe Pinto, Han Li, Tianzhen Hong, Mohamed Ouf, Alfonso Capozzoli, and Zoltan Nagy. 2023a, November. “CityLearn v2: An OpenAI Gym environment for demand response control benchmarking in grid-interactive communities.” *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, BuildSys ’23. New York, NY, USA: Association for Computing Machinery, 274–275.

Nweye, Kingsley, Kathryn Kaspar, Giacomo Buscemi, Giuseppe Pinto, Han Li, Tianzhen Hong, Mohamed Ouf, Alfonso Capozzoli, and Zoltan Nagy. 2023b, September. “A framework for the design of representative neighborhoods for energy flexibility assessment in CityLearn.” *Proceedings of building simulation 2023: 18th conference of IBPSA*, Volume 18 of *Building simulation*. Shanghai, China: IBPSA, 3351–3358. ISSN: 2522-2708 tex.organisation: IBPSA.

Nweye, Kingsley, Bo Liu, Peter Stone, and Zoltan Nagy. 2022a. “Real-world challenges for multi-agent reinforcement learning in grid-interactive buildings.” *Energy and AI* 10:100202.

Nweye, Kingsley, Zoltan Nagy, Sharada Mohanty, Dipam Chakraborty, Siva Sankaranarayanan, Tianzhen Hong, Sourav Dey, Gregor Henze, Jan Drgona, Fangquan Lin, Wei Jiang, Hanwei Zhang, Zhongkai Yi, Jihai Zhang, Cheng Yang, Matthew Motoki, Sorapong Khongnawang, Michael Ibrahim, Abilmansur Zhumabekov, Daniel May, Zhihu Yang, Xiaozhuang Song, Han Zhang, Xiaoning Dong, Shun Zheng, and Jiang Bian. 2022b, August. “The CityLearn Challenge 2022: Overview, Results, and Lessons Learned.” *Proceedings of the NeurIPS 2022 Competitions Track*. PMLR, 85–103. ISSN: 2640-3498.

Nweye, Kingsley, Siva Sankaranarayanan, and Zoltan Nagy. 2023. “MERLIN: Multi-agent offline and transfer learning for occupant-centric operation of grid-interactive communities.” *Applied Energy* 346 (September): 121323.

NYSERDA. 2022, July. NYSERDA RTEM Hackathon Demo Day.

Park, June Young, and Zoltan Nagy. 2020, June. “HVACLearn: A reinforcement learning based occupant-centric control for thermostat set-points.” *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*, e-Energy ’20. New York, NY, USA: Association for Computing Machinery, 434–437.

Pinto, Giuseppe, Davide Deltetto, and Alfonso Capozzoli. 2021. “Data-driven district energy management with surrogate models and deep reinforcement learning.” *Applied Energy* 304 (December): 117642.

Raffin, Antonin, Ashley Hill, Adam Gleave, Anssi Kanervisto, Maximilian Ernestus, and Noah Dormann. 2021. “Stable-Baselines3: Reliable Reinforcement Learning Implementations.” *Journal of Machine Learning Research* 22 (268): 1–8.

Scharnhorst, Paul, Baptiste Schubnel, Carlos Fernández Bandera, Jaume Salom, Paolo Taddeo, Max Boegli, Tomasz Gorecki, Yves Stauffer, Antonis Peppas, and Chrysa Politi. 2021. “Energym: A building model library for controller benchmarking.” *Applied Sciences* 11 (8): 3518. Publisher: MDPI.

U.S. Department of Energy. 2021. “Next-Generation Grid Technologies.” Technical Report, U.S. Department of Energy.

U.S. Energy Information Administration. 2020 Residential Energy Consumption Survey.

Vázquez-Canteli, José R., and Zoltán Nagy. 2019. “Reinforcement learning for demand response: A review of algorithms and modeling techniques.” *Applied Energy* 235 (February): 1072–1089.

Vázquez-Canteli, J.R. José R., Jérôme Kämpf, Gregor Henze, and Zoltan Nagy. 2019. “CityLearn v1.0: An OpenAI gym environment for demand response with deep reinforcement learning.” *BuildSys 2019 - Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, pp. 356–357. ISBN: 9781450370059.

Wei, Mingjun, Wenqi Jia, Yangyang Fu, Zhiyao Yang, and Zheng O’Neill. 2023, February. *Energy Management and Control System for a PV-Battery System to Improve Residential Building Resiliency Under Extreme Weather Conditions*.

Wilson, Eric J. H., Andrew Parker, Anthony Fontanini, Elaina Present, Janet L. Reyna, Rajendra Adhikari, Carlo Bianchi, Christopher CaraDonna, Matthew Dahlhausen, Janghyun Kim, Amy LeBar, Lixi Liu, Marlena Praprost, Liang Zhang, Peter DeWitt, Noel Merket, Andrew Speake, Tianzhen Hong, Han Li, Natalie Mims Frick, Zhe Wang, Aileen Blair, Henry Horsey, David Roberts, Kim Trenbath, and Oluwatobi Adekanye. 2022. “End-Use Load Profiles for the U.S. Building Stock: Methodology and Results of Model Calibration, Validation, and Uncertainty Quantification.” March.